

VOLUME 38

OCTOBER 1960

NUMBER 10

# Canadian Journal of Physics

**Editor:** H. E. DUCKWORTH

**Associate Editors:**

L. G. ELLIOTT, *Atomic Energy of Canada, Ltd., Chalk River*

J. S. FOSTER, *McGill University*

G. HERZBERG, *National Research Council of Canada*

L. LEPRINCE-RINGUET, *Ecole Polytechnique, Paris*

B. W. SARGENT, *Queen's University*

G. M. VOLKOFF, *University of British Columbia*

W. H. WATSON, *University of Toronto*

G. A. WOONTON, *McGill University*

**Published by** THE NATIONAL RESEARCH COUNCIL

OTTAWA

CANADA

## CANADIAN JOURNAL OF PHYSICS

Under the authority of the Chairman of the Committee of the Privy Council on Scientific and Industrial Research, the National Research Council issues THE CANADIAN JOURNAL OF PHYSICS and five other journals devoted to the publication, in English or French, of the results of original scientific research. Matters of general policy concerning these journals are the responsibility of a joint Editorial Board consisting of: members representing the National Research Council of Canada; the Editors of the Journals; and members representing the Royal Society of Canada and four other scientific societies.

### EDITORIAL BOARD

#### Representatives of the National Research Council

I. McT. Cowan (Chairman), *University of British Columbia*  
Léo Marion, *National Research Council*

H. G. Thode, *McMaster University*  
D. L. Thomson, *McGill University*

#### Editors of the Journals

D. L. Bailey, *University of Toronto*  
T. W. M. Cameron, *Macdonald College*  
F. E. Chase, *Ontario Agricultural College*  
H. E. Duckworth, *McMaster University*

Léo Marion, *National Research Council*  
J. F. Morgan, *Department of National Health and Welfare, Ottawa*  
J. A. F. Stevenson, *University of Western Ontario*

#### Representatives of Societies

D. L. Bailey, *University of Toronto*  
Royal Society of Canada  
T. W. M. Cameron, *Macdonald College*  
Royal Society of Canada  
H. E. Duckworth, *McMaster University*  
Royal Society of Canada  
Canadian Association of Physicists  
P. R. Gendron, *University of Ottawa*  
Chemical Institute of Canada

D. J. Le Roy, *University of Toronto*  
Royal Society of Canada  
J. F. Morgan, *Department of National Health and Welfare, Ottawa*  
Canadian Biochemical Society  
R. G. E. Murray, *University of Western Ontario*  
Canadian Society of Microbiologists  
J. A. F. Stevenson, *University of Western Ontario*  
Canadian Physiological Society

#### Ex officio

Léo Marion (Editor-in-Chief), *National Research Council*  
J. B. Marshall (Administration and Awards), *National Research Council*

*Manuscripts* for publication should be submitted to Dr. H. E. Duckworth, Editor, Canadian Journal of Physics, Hamilton College, McMaster University, Hamilton, Ontario.

For instructions on preparation of copy, see **NOTES TO CONTRIBUTORS** (back cover).

*Proof, correspondence concerning proof, and orders for reprints* should be sent to the Manager, Editorial Office (Research Journals), Division of Administration and Awards, National Research Council, Ottawa 2, Canada.

*Subscriptions, renewals, requests for single or back numbers, and all remittances* should be sent to Division of Administration and Awards, National Research Council, Ottawa 2, Canada. Remittances should be made payable to the Receiver General of Canada, credit National Research Council.

The journals published, frequency of publication, and subscription prices are:

Canadian Journal of Biochemistry and Physiology	Monthly	\$9.00 a year
Canadian Journal of Botany	Bimonthly	\$6.00 a year
Canadian Journal of Chemistry	Monthly	\$12.00 a year
Canadian Journal of Microbiology	Bimonthly	\$6.00 a year
Canadian Journal of Physics	Monthly	\$9.00 a year
Canadian Journal of Zoology	Bimonthly	\$5.00 a year

The price of regular single numbers of all journals is \$2.00.







# Canadian Journal of Physics

Issued by THE NATIONAL RESEARCH COUNCIL OF CANADA

VOLUME 38

OCTOBER 1960

NUMBER 10

## AN ELECTROMAGNETIC DIFFRACTION PROBLEM INVOLVING UNIDIRECTIONALLY CONDUCTING SURFACES<sup>1</sup>

R. A. HURD

### ABSTRACT

The exact value of the electromagnetic field scattered by a unidirectionally conducting surface under plane wave excitation is obtained. The surface, which is an entire plane, is the junction of two unidirectionally conducting half-planes whose conductivity directions are inclined at an arbitrary angle to each other, and to the boundary line.

### INTRODUCTION

By a unidirectionally conducting (u.c.) surface is understood a surface which is perfectly conducting in one direction, and is perfectly insulating in the orthogonal direction. An array of parallel conducting wires in the limit of thickness and spacing tending to zero is an example.

Diffraction problems involving u.c. surfaces have received attention recently from several authors. Toraldo di Francia (1956a) has solved the problem of diffraction of plane waves by a small u.c. disk, while Karp (1957), Radlow (1959), and Hurd (1960) have considered diffraction by a u.c. half-plane.

In this paper we solve the following problem. A plane electromagnetic wave falls upon the surface  $z = 0$ . This plane is divided into two regions by the line  $x = 0$ . On the half-plane  $x > 0$ ,  $z = 0$  a rotated co-ordinate system  $(\xi_1, \eta_1)$  is set up, where

$$\begin{aligned}\xi_1 &= x \cos \alpha_1 + y \sin \alpha_1, \\ \eta_1 &= -x \sin \alpha_1 + y \cos \alpha_1.\end{aligned}$$

This half-plane is assumed to be u.c. in the  $\xi_1$  direction.

In  $x < 0$ ,  $z = 0$  a second co-ordinate system  $(\xi_2, \eta_2)$  is introduced, given by

$$\begin{aligned}\xi_2 &= x \cos \alpha_2 + y \sin \alpha_2, \\ \eta_2 &= -x \sin \alpha_2 + y \cos \alpha_2.\end{aligned}$$

This half-plane is taken to be u.c. in the  $\xi_2$  direction. Both  $\alpha_1$  and  $\alpha_2$  are measured counter-clockwise from the  $x$ -axis.

It is required to find the scattered field generated by this surface.

<sup>1</sup>Manuscript received June 3, 1960.

Contribution from the Division of Radio and Electrical Engineering, National Research Council, Ottawa, Canada.

Issued as N.R.C. No. 5931.

## SOLUTION OF THE PROBLEM

Let the incident field be given by

$$(1a) \quad \mathbf{E}^{(i)} = \mathbf{A} \exp (i\mathbf{k} \cdot \mathbf{r} - i\omega t),$$

$$(1b) \quad \mathbf{H}^{(i)} = (kz_0)^{-1}(\mathbf{k} \times \mathbf{A}) \exp (i\mathbf{k} \cdot \mathbf{r} - i\omega t),$$

where

$$\mathbf{k} = \{k_1, k_2, k_3\},$$

$$\mathbf{r} = \{x, y, z\},$$

$$k = (k_1^2 + k_2^2 + k_3^2)^{\frac{1}{2}},$$

$$z_0 = (\mu_0/\epsilon_0)^{\frac{1}{2}}.$$

We denote the total fields by  $\mathbf{E}$ ,  $\mathbf{H}$ , and the scattered fields by  $\mathbf{E}^{(s)}$ ,  $\mathbf{H}^{(s)}$ ; thus  $\mathbf{E} = \mathbf{E}^{(i)} + \mathbf{E}^{(s)}$ ,  $\mathbf{H} = \mathbf{H}^{(i)} + \mathbf{H}^{(s)}$ .

The boundary conditions which must be satisfied on the screen  $x > 0$ ,  $z = 0$  are

$$(2a) \quad E_{\xi_1} = 0 \text{ on the screen,}$$

$$(2b) \quad H_{\xi_1} \text{ continuous across the screen;}$$

and for  $x < 0$ ,  $z = 0$

$$(2c) \quad E_{\xi_2} = 0 \text{ on the screen,}$$

$$(2d) \quad H_{\xi_2} \text{ continuous across the screen.}$$

Owing to the planar nature of the problem, the scattered fields must obey the relations

$$(3a) \quad H_{x,y}^{(s)}(z) = -H_{x,y}^{(s)}(-z),$$

$$(3b) \quad E_{x,y}^{(s)}(z) = E_{x,y}^{(s)}(-z),$$

$$(3c) \quad H_z^{(s)}(z) = H_z^{(s)}(-z),$$

$$(3d) \quad E_z^{(s)}(z) = -E_z^{(s)}(-z).$$

A proof of eqs. (3) has been given by Bouwkamp (1954).

Henceforth, we consider fields in the region  $z \geq 0$  only. Once the solution here is known, the fields in  $z < 0$  may be obtained from eqs. (3).

Taking eqs. (3) into account, we are able to express boundary conditions (2b) and (2d) in the following way:

$$(2b)' \quad H_{\xi_1}^{(s)} = 0, \quad x > 0,$$

$$(2d)' \quad H_{\xi_2}^{(s)} = 0, \quad x < 0.$$

We look for a solution in the form

$$(4a) \quad \mathbf{E}^{(s)} = \nabla \times \nabla \times \mathbf{\Pi}^{(s)} + i\omega\mu_0 \nabla \times \mathbf{\Pi}^{(s)*},$$

$$(4b) \quad \mathbf{H}^{(s)} = \nabla \times \nabla \times \mathbf{\Pi}^{(s)*} - i\omega\epsilon_0 \nabla \times \mathbf{\Pi}^{(s)},$$

with

$$(5a) \quad \Pi^{(s)} = \{0, \Pi_y^{(s)}, 0\},$$

$$(5b) \quad \Pi^{(s)*} = \{0, \Pi_y^{(s)*}, 0\},$$

and where

$$(6a) \quad \Pi_y^{(s)} = \int_{-\infty+ia}^{\infty+ia} f(w) e^{i(wx+\beta z)} dw + F e^{i(k_1 x + k_3 z)},$$

$$(6b) \quad \Pi_y^{(s)*} = \int_{-\infty+ia}^{\infty+ia} g(w) e^{i(wx+\beta z)} dw + G e^{i(k_1 x + k_3 z)}.$$

In eqs. (6) a multiplicative factor  $\exp(ik_2 y - i\omega t)$  has been omitted. The amplitude functions  $f$ ,  $F$ ,  $g$ ,  $G$  are at present unknown. Also

$$\beta^2 = \kappa^2 - w^2,$$

$$\kappa^2 = k^2 - k_2^2.$$

We choose the branch of  $\beta$  which tends to  $\kappa$  as  $w \rightarrow 0$ . For the time being,  $\kappa$  is supposed to be complex  $\kappa = p_2 + iq_2$ ,  $q_2 > 0$ . We also assume that  $k_1 = p_1 + iq_1$ , where  $p_1$  and  $q_1$  have the same sign, though they are not necessarily positive. Since  $k_1, k_2, k_3$  are proportional to  $k$ ,  $k_1$  must be proportional to  $\kappa$ , the constant of proportionality being real. Using this fact, it is easy to show that  $|q_1| \leq q_2$ . The path of integration in eqs. (6) is taken to run below the point  $w = k_1$ ; that is,  $q_1 > a \geq -q_2$ .

From eqs. (4), (5), and (6) we compute the  $\xi_1$  and  $\xi_2$  field components.

$$(7a) \quad H_{\xi_{1,2}}^{(s)} = \int_{-\infty+ia}^{\infty+ia} \{ \kappa^2 g(w) \sin \alpha_{1,2} - [\omega \epsilon_0 \beta f(w) + k_2 w g(w)] \cos \alpha_{1,2} \} e^{i(wx+\beta z)} dw \\ + \{ \kappa^2 G \sin \alpha_{1,2} - [\omega \epsilon_0 k_3 F + k_1 k_2 G] \cos \alpha_{1,2} \} e^{i(k_1 x + k_3 z)},$$

$$(7b) \quad E_{\xi_{1,2}}^{(s)} = \int_{-\infty+ia}^{\infty+ia} \{ \kappa^2 f(w) \sin \alpha_{1,2} + [\omega \mu_0 \beta g(w) - k_2 w f(w)] \cos \alpha_{1,2} \} e^{i(wx+\beta z)} dw \\ + \{ \kappa^2 F \sin \alpha_{1,2} + [\omega \mu_0 k_3 G - k_1 k_2 F] \cos \alpha_{1,2} \} e^{i(k_1 x + k_3 z)}.$$

For  $x < 0$ ,  $z = 0$  the integrals in eqs. (7) can contribute no terms of the form  $e^{i(k_1 x + k_3 z)}$  since the path of integration is drawn below  $w = k_1$ , and since closure of the contour must be carried out in the lower half  $w$ -plane. Hence in order to satisfy  $H_{\xi_2}^{(s)} = 0$  for  $x < 0$  we must have

$$(8a) \quad -(\omega \epsilon_0 k_3 F + k_1 k_2 G) \cos \alpha_2 + \kappa^2 G \sin \alpha_2 = 0.$$

Also, to make  $E_{\xi_2} = 0$  for  $x < 0$

$$(8b) \quad (\omega \mu_0 k_3 G - k_1 k_2 F) \cos \alpha_2 + \kappa^2 F \sin \alpha_2 = -A_{\xi_2}.$$

And thus, if  $P_2(k_1) = (k_1 k_2 \cos \alpha_2 - \kappa^2 \sin \alpha_2)^2 + k^2 k_3^2 \cos^2 \alpha_2$ ,

$$(9a) \quad F = (k_1 k_2 \cos \alpha_2 - \kappa^2 \sin \alpha_2) A_{\xi_2} / P_2(k_1),$$

$$(9b) \quad G = -\omega \epsilon_0 k_3 \cos \alpha_2 A_{\xi_2} / P_2(k_1).$$

In order that the integral parts of  $H_{\xi_2}^{(s)}$  and  $E_{\xi_2}^{(s)}$  contribute nothing when  $x < 0$ ,  $z = 0$ , we require that

$$(10a) \quad -[\omega\epsilon_0\beta f(w) + k_2wg(w)] \cos \alpha_2 + \kappa^2g(w) \sin \alpha_2 = L_1(w),$$

$$(10b) \quad [\omega\mu_0\beta g(w) - k_2wf(w)] \cos \alpha_2 + \kappa^2f(w) \sin \alpha_2 = L_2(w),$$

where  $L_1(w)$  and  $L_2(w)$  are regular functions in the lower part of the  $w = u + iv$  plane,  $v \leq a + \delta$ ,  $\delta > 0$ .

In the region  $x > 0$ ,  $z = 0$ , the boundary conditions  $H_{\xi_1}^{(s)} = 0$ ,  $E_{\xi_1} = 0$  can be satisfied if

$$(11a) \quad -[\omega\epsilon_0\beta f(w) + k_2wg(w)] \cos \alpha_1 + \kappa^2g(w) \sin \alpha_1 = U_1(w)/(w - k_1),$$

$$(11b) \quad [\omega\mu_0\beta g(w) - k_2wf(w)] \cos \alpha_1 + \kappa^2f(w) \sin \alpha_1 = U_2(w)/(w - k_1),$$

where  $U_1(w)$  and  $U_2(w)$  are regular for  $v \geq a - \delta$ ; and providing that  $U_1(k_1)$  and  $U_2(k_1)$  are chosen so that

$$(12a) \quad 2\pi i U_1(k_1) - [\omega\epsilon_0k_3F + k_1k_2G] \cos \alpha_1 + \kappa^2G \sin \alpha_1 = 0,$$

$$(12b) \quad 2\pi i U_2(k_1) + [\omega\mu_0k_3G - k_1k_2F] \cos \alpha_1 + \kappa^2F \sin \alpha_1 = -A_{\xi_1}.$$

We now eliminate  $f(w)$  and  $g(w)$  from eqs. (10) and (11). This is accomplished by solving eqs. (10) for  $f(w)$  and  $g(w)$  and substituting the values obtained in eqs. (11). There results two equations in the remaining four unknowns,

$$(13a) \quad U_1(w)P_2(w)(w - k_1)^{-1} = \omega\epsilon_0\beta\kappa^2L_2(w) \sin(\alpha_1 - \alpha_2) + Q(w)L_1(w),$$

$$(13b) \quad U_2(w)P_2(w)(w - k_1)^{-1} = Q(w)L_2(w) - \omega\mu_0\beta\kappa^2L_1(w) \sin(\alpha_1 - \alpha_2),$$

where

$$(14) \quad P_2(w) = (\kappa^2 \sin \alpha_2 - k_2w \cos \alpha_2)^2 + k^2\beta^2 \cos^2 \alpha_2,$$

$$(15) \quad Q(w) = (\kappa^2 \sin \alpha_1 - k_2w \cos \alpha_1)(\kappa^2 \sin \alpha_2 - k_2w \cos \alpha_2) + k^2\beta^2 \cos \alpha_1 \cos \alpha_2.$$

If eq. (13a) is multiplied by  $\pm iz_0$  and then is added to eq. (13b), there results

$$(16) \quad \frac{U_2(w) \pm iz_0 U_1(w)}{w - k_1} = \frac{[L_2(w) \pm iz_0 L_1(w)][Q(w) \pm i\beta\sigma]}{P_2(w)},$$

where  $\sigma = k\kappa^2 \sin(\alpha_1 - \alpha_2)$ .

We put

$$(17a) \quad J(w) = [Q(w) + i\beta\sigma]/P_2(w),$$

$$(17b) \quad K(w) = [Q(w) - i\beta\sigma]/P_2(w),$$

and suppose that  $J$  and  $K$  can be factored into upper and lower functions, viz.

$$J(w) = J_-(w)/J_+(w),$$

$$K(w) = K_-(w)/K_+(w),$$

where  $J_-(w)$  is regular and non-zero for  $v < q_2$ , and  $J_+(w)$  is regular and non-zero for  $v > -q_2$ ; similarly for  $K_-(w)$  and  $K_+(w)$ .

Then, if we introduce the abbreviations

$$U_3(w) = U_2(w) + iz_0 U_1(w),$$

$$U_4(w) = U_2(w) - iz_0 U_1(w),$$

with analogous definitions for  $L_3(w)$  and  $L_4(w)$ , eqs. (16) can be written as

$$(18a) \quad \frac{J_+(w)U_3(w) - J_+(k_1)U_3(k_1)}{w - k_1} = J_-(w)L_3(w) - \frac{J_+(k_1)U_3(k_1)}{w - k_1},$$

$$(18b) \quad \frac{K_+(w)U_4(w) - K_+(k_1)U_4(k_1)}{w - k_1} = K_-(w)L_4(w) - \frac{K_+(k_1)U_4(k_1)}{w - k_1}.$$

The left sides of eqs. (18) are regular in  $v \geq a - \delta$ ; the right sides are regular in  $v \leq a + \delta$ . Hence by analytic continuation each must be equal to functions  $R(w)$  regular throughout the entire  $w$ -plane.

Now in order that the integrals defining  $E_{\xi_1,2}^{(s)}$  and  $H_{\xi_1,2}^{(s)}$  converge when  $z = 0$ , we must have

$$U_1(w), U_2(w) = O(w^{1-\epsilon}),$$

$$L_1(w), L_2(w) = O(w^{-\epsilon}),$$

for  $|w| \rightarrow \infty$ , where  $\epsilon > 0$ . And since, as will be shown, the factors of  $J(w)$  and  $K(w)$  are  $O(1)$  as  $|w| \rightarrow \infty$ ,\* we then have that  $R(w) = O(w^{-\epsilon})$  for  $|w| \rightarrow \infty$ . Hence by Liouville's theorem  $R(w) = 0$ .

The solution to eqs. (18) is

$$(19a) \quad U_3(w) = U_3(k_1)J_+(k_1)/J_+(w),$$

$$(19b) \quad U_4(w) = U_4(k_1)K_+(k_1)/K_+(w),$$

$$(19c) \quad L_3(w) = U_3(k_1)J_+(k_1)/[(w - k_1)J_-(w)],$$

$$(19d) \quad L_4(w) = U_4(k_1)K_+(k_1)/[(w - k_1)K_-(w)].$$

$U_3(k_1)$  and  $U_4(k_1)$  are obtained from eqs. (9) and (12); thus

$$(20a) \quad U_3(k_1) = (i/2\pi)[A_{\xi_1} - J(k_1)A_{\xi_2}],$$

$$(20b) \quad U_4(k_1) = (i/2\pi)[A_{\xi_1} - K(k_1)A_{\xi_2}].$$

The functions  $f(w)$  and  $g(w)$  can be found from the relations

$$(21a) \quad f(w) = (k/\sigma)[U_2(w)(w - k_1)^{-1} \cos \alpha_2 - L_2(w) \cos \alpha_1],$$

$$(21b) \quad g(w) = (k/\sigma)[U_1(w)(w - k_1)^{-1} \cos \alpha_2 - L_1(w) \cos \alpha_1],$$

which follow from eqs. (10) and (11).

\*An exception occurs if either  $\alpha_1$  or  $\alpha_2 = \pi/2$ , but the conclusion  $R(w) = 0$  is unchanged.

If we introduce the integrals

$$(22a) \quad I_1(x, z, p, q) = \int_{-\infty+ia}^{\infty+ia} \left[ \frac{p}{J_+(w)} + \frac{q}{K_+(w)} \right] e^{i(\alpha x + \beta z)} \frac{dw}{w - k_1},$$

$$(22b) \quad I_2(x, z, p, q) = \int_{-\infty+ia}^{\infty+ia} \left[ \frac{p}{J_-(w)} + \frac{q}{K_-(w)} \right] e^{i(\alpha x + \beta z)} \frac{dw}{w - k_1},$$

where

$$p = (i/2\pi)[A_{\xi_1}J_+(k_1) - A_{\xi_2}J_-(k_1)],$$

$$q = (i/2\pi)[A_{\xi_1}K_+(k_1) - A_{\xi_2}K_-(k_1)],$$

then the complete solution is given by

$$(23a) \quad \Pi_y^{(s)} = (k_1 k_2 \cos \alpha_2 - \kappa^2 \sin \alpha_2) P_2^{-1}(k_1) A_{\xi_2} e^{i(k_1 x + k_2 z)} \\ + (k/2\sigma)[I_1(x, z, p, q) \cos \alpha_2 - I_2(x, z, p, q) \cos \alpha_1],$$

$$(23b) \quad \Pi_y^{(s)*} = -k k_3 z_0^{-1} P_2^{-1}(k_1) A_{\xi_2} \cos \alpha_2 e^{i(k_1 x + k_2 z)} \\ - (ik/2\sigma z_0)[I_1(x, z, p, -q) \cos \alpha_2 - I_2(x, z, p, -q) \cos \alpha_1].$$

The tangential field components are readily obtained from eqs. (7), (10), (11), (19).

$$(24a) \quad E_{\xi_1}^{(s)} = -Q(k_1) P_2^{-1}(k_1) A_{\xi_2} e^{i(k_1 x + k_2 z)} + \frac{1}{2} I_1(x, z, p, q),$$

$$(24b) \quad E_{\xi_2}^{(s)} = -A_{\xi_2} e^{i(k_1 x + k_2 z)} + \frac{1}{2} I_2(x, z, p, q),$$

$$(24c) \quad H_{\xi_1}^{(s)} = -\sigma k_3 z_0^{-1} P_2^{-1}(k_1) A_{\xi_2} e^{i(k_1 x + k_2 z)} - (i/2z_0) I_1(x, z, p, -q),$$

$$(24d) \quad H_{\xi_2}^{(s)} = -(i/2z_0) I_2(x, z, p, -q).$$

#### FACTORING $J(w)$ AND $K(w)$

We consider  $J(w)$  first. Referring to eqs. (14) and (17a) we see that, since  $P_2(w)$  is a quadratic in  $w$ , the only term involving difficulty is  $Q + i\beta\sigma$ , since this expression has branch points at  $w = \pm\kappa$ .

We choose the branch cuts of  $\beta$  to be radial lines through  $\pm\kappa$ , as shown in Fig. 1. We divide the  $w$ -plane near the branch cuts into four regions which are bounded by the branch cuts and by the lines  $\text{Im}(\beta) = 0$  and  $\text{Re}(\beta) = 0$  (Fig. 1). In regions (1) and (3)  $\text{Im}(\beta) > 0$ , while in regions (2) and (4)  $\text{Im}(\beta) < 0$ .

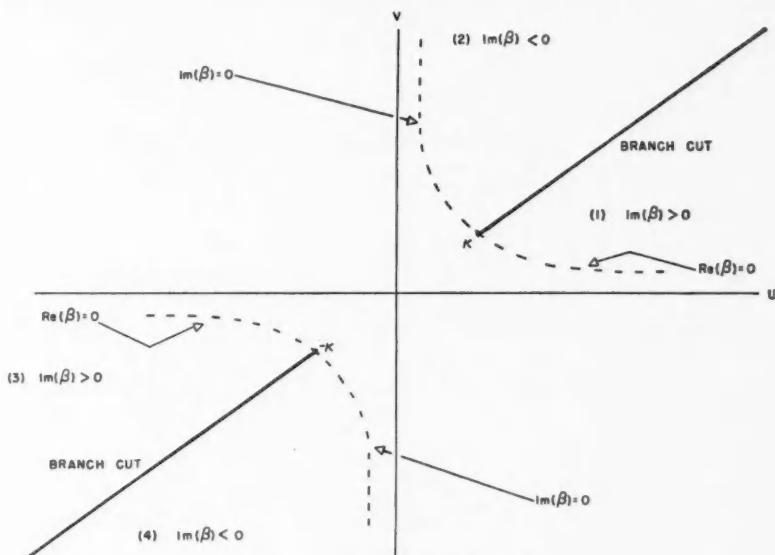
Now  $Q + i\beta\sigma$  has four zeros, which are given by

$$(25a) \quad a_{1,2} = k \sec \alpha_{1,2} - k_2 \tan \alpha_{1,2},$$

$$(25b) \quad b_{1,2} = -k \sec \alpha_{1,2} - k_2 \tan \alpha_{1,2}.$$

Bearing in mind what has been said about the proportionality of  $k$ ,  $\kappa$ ,  $k_1$ ,  $k_2$ ,  $k_3$ , it is not difficult to show that  $a_1$  and  $a_2$  lie on the branch cut in the upper half-plane, while  $b_1$  and  $b_2$  lie in the branch cut in the lower half-plane.



FIG. 1. The branch cuts and regions of the  $w$ -plane.

To find which regions of the  $w$ -plane the zeros belong to, we first calculate the values of  $\beta$  as  $w$  tends to these points.

$$(26a) \quad \beta(a_{1,2}) = \pm i(k \tan \alpha_{1,2} - k_2 \sec \alpha_{1,2}),$$

$$(26b) \quad \beta(b_{1,2}) = \pm i(k \tan \alpha_{1,2} + k_2 \sec \alpha_{1,2}).$$

In accordance with Fig. 1, the upper sign in eq. (26a) is to be taken if  $\text{Re}(k \sin \alpha_{1,2}) > \text{Re}(k_2)$  and if  $w \rightarrow {}_1a_{1,2}$ ; or if  $\text{Re}(k \sin \alpha_{1,2}) < \text{Re}(k_2)$  and  $w \rightarrow {}_2a_{1,2}$ . Otherwise the lower sign is to be used. (The notation  ${}_1a_{1,2}$  means the limit as  $w \rightarrow a_{1,2}$  from points within region (1).) Analogous remarks apply to eq. (26b).

By direct substitution we find that the values of  $Q(w)$  at the zeros are

$$(27a) \quad Q(a_1) = -\sigma(k \tan \alpha_1 - k_2 \sec \alpha_1),$$

$$(27b) \quad Q(a_2) = \sigma(k \tan \alpha_2 - k_2 \sec \alpha_2),$$

$$(27c) \quad Q(b_1) = -\sigma(k \tan \alpha_1 + k_2 \sec \alpha_1),$$

$$(27d) \quad Q(b_2) = \sigma(k \tan \alpha_2 + k_2 \sec \alpha_2).$$

Hence, using eqs. (26) and the remarks following eqs. (26), we find that the arrangement of the zeros of  $Q(w) + i\beta\sigma$  is as given in Table I.

TABLE I  
The zeros of  $Q(w) + i\beta\sigma$

Real part of:	$k \sin \alpha_1 > k_2$		$k_2 > k \sin \alpha_1 > -k_2$		$-k_2 > k \sin \alpha_1$	
$k \sin \alpha_2$ $> k_2$	${}_2a_1,$ ${}_4b_1,$	${}_1a_2$ ${}_3b_2$	${}_1a_1,$ ${}_4b_1,$	${}_1a_2$ ${}_3b_2$	${}_1a_1,$ ${}_3b_1,$	${}_1a_2$ ${}_3b_2$
$k_2 >$ $k \sin \alpha_2 >$ $-k_2$	${}_2a_1,$ ${}_4b_1,$	${}_2a_2$ ${}_3b_2$	${}_1a_1,$ ${}_4b_1,$	${}_2a_2$ ${}_3b_2$	${}_1a_1,$ ${}_3b_1,$	${}_2a_2$ ${}_3b_2$
$-k_2 >$ $k \sin \alpha_2$	${}_2a_1,$ ${}_4b_1,$	${}_2a_2$ ${}_4b_2$	${}_1a_1,$ ${}_4b_1,$	${}_2a_2$ ${}_4b_2$	${}_1a_1,$ ${}_3b_1,$	${}_2a_2$ ${}_4b_2$

We now expand the logarithmic derivative of  $Q(w) + i\beta\sigma$  in partial fractions

$$(28) \quad \frac{(Q+i\beta\sigma)'}{Q+i\beta\sigma} = \frac{A_1}{w-a_1} + \frac{A_1'}{\beta(w-a_1)} + \frac{A_2}{w-a_2} + \frac{A_2'}{\beta(w-a_2)} \\ + \frac{B_1}{w-b_1} + \frac{B_1'}{\beta(w-b_1)} + \frac{B_2}{w-b_2} + \frac{B_2'}{\beta(w-b_2)},$$

To determine the coefficients  $A_1, A_1'$ , we multiply by  $w-a_1$  and go to the limit  $w \rightarrow a_1$  in regions (1) and (2). If  $\text{Re}(k \sin \alpha_1) > \text{Re}(k_2)$  we have

$$(29a) \quad 0 = A_1 + A_1'/\beta({}_1a_1),$$

$$(29b) \quad 1 = A_1 + A_1'/\beta({}_2a_1),$$

from which  $A_1$  and  $A_1'$  can be calculated. On the other hand, if  $\text{Re}(k \sin \alpha_1) < \text{Re}(k_2)$ , the following equations result:

$$(30a) \quad 1 = A_1 + A_1'/\beta({}_1a_1),$$

$$(30b) \quad 0 = A_1 + A_1'/\beta({}_2a_1).$$

The remaining coefficients are determined in like manner. It will be observed that  $\beta({}_1a_1)$  in eqs. (29) equals  $\beta({}_2a_1)$  in eqs. (30), and vice versa, so that eqs. (29) and (30) are equivalent. Hence a single expression, which holds for all cases, is obtained for the logarithmic derivative:

$$(31) \quad \frac{2(Q+i\beta\sigma)'}{Q+i\beta\sigma} = \frac{1}{w-a_1} \left[ 1 - \frac{\beta(a_1)}{\beta} \right] + \frac{1}{w-a_2} \left[ 1 + \frac{\beta(a_2)}{\beta} \right] \\ + \frac{1}{w-b_1} \left[ 1 - \frac{\beta(b_1)}{\beta} \right] + \frac{1}{w-b_2} \left[ 1 + \frac{\beta(b_2)}{\beta} \right]$$

providing that the following definitions are imposed

$$(32a) \quad \beta(a_{1,2}) = i(k \tan \alpha_{1,2} - k_2 \sec \alpha_{1,2}),$$

$$(32b) \quad \beta(b_{1,2}) = i(k \tan \alpha_{1,2} + k_2 \sec \alpha_{1,2}).$$

\*That an expansion of this kind is possible can be shown as follows:

$$\frac{(Q+i\beta\sigma)'}{Q+i\beta\sigma} = \frac{(Q' - i w \sigma / \beta)(Q - i\beta\sigma)}{Q^2 + \beta^2 \sigma^2} = \frac{[QQ' - w\sigma^2] - (i\sigma/\beta)[wQ + \beta^2 Q']}{Q^2 + \beta^2 \sigma^2}.$$

The expressions in square brackets are cubic, while the denominator is quartic in  $w$ , hence each quotient can be expanded in partial fractions.

Now  $1/\beta$  can be written as

$$(33) \quad \frac{1}{\beta} = \frac{1}{\pi} \left[ \frac{\pi/2 + \sin^{-1} w/\kappa}{\beta} + \frac{\pi/2 - \sin^{-1} w/\kappa}{\beta} \right],$$

where the branch of  $\sin^{-1} w/\kappa$  intended is that for which

$$-\pi/2 \leq \operatorname{Re}(\sin^{-1} w/\kappa) \leq \pi/2.$$

The first term of eq. (33) is regular for  $v < q_2$ , the second term is regular when  $v > -q_2$ . Hence

$$(34) \quad \frac{1}{(w-a_1)} = \frac{1}{\pi(w-a_1)} \left[ \frac{\pi/2 + \sin^{-1} w/\kappa}{\beta} + \frac{\pi/2 - \sin^{-1} a_1/\kappa}{\beta(a_1)} \right] \\ + \frac{1}{\pi(w-a_1)} \left[ \frac{\pi/2 - \sin^{-1} w/\kappa}{\beta} - \frac{\pi/2 - \sin^{-1} a_1/\kappa}{\beta(a_1)} \right].$$

The first and second terms of eq. (34) are regular for  $v < q_2$  and  $v > -q_2$ , respectively. The interpretation of  $\sin^{-1} a_1/\kappa$  in eq. (34) is

$$\sin^{-1} a_1/\kappa = \sin^{-1} {}_1a_1/\kappa \quad \text{for } \operatorname{Re}(k \sin \alpha_1) > \operatorname{Re}(k_2), \\ = \sin^{-1} {}_2a_1/\kappa \quad \text{for } \operatorname{Re}(k \sin \alpha_1) < \operatorname{Re}(k_2);$$

and  $\beta(a_1)$  is defined by eq. (32a).

Now if  $Q+i\beta\sigma = (Q+i\beta\sigma)_-/(Q+i\beta\sigma)_+$ , then

$$(35) \quad \frac{(Q+i\beta\sigma)'_+}{(Q+i\beta\sigma)_-} = \frac{(Q+i\beta\sigma)'_-}{(Q+i\beta\sigma)_-} - \frac{(Q+i\beta\sigma)'_+}{(Q+i\beta\sigma)_+},$$

where the  $(-)$  and  $(+)$  subscripts indicate regularity in the lower and upper half-planes, respectively. So that, from eqs. (31) and (34) we have, after some rearrangement

$$(36a) \quad \frac{(Q+i\beta\sigma)'_-}{(Q+i\beta\sigma)_-} = \frac{1}{w-a_2} - \frac{\beta(a_1)}{2\pi(w-a_1)} [F(w) - F(a_1)] + \frac{\beta(a_2)}{2\pi(w-a_2)} \\ \times [F(w) - F(a_2)] - \frac{\beta(b_1)}{2\pi(w-b_1)} [F(w) - F(b_1)] + \frac{\beta(b_2)}{2\pi(w-b_2)} \\ [F(w) - F(b_2)],$$

$$(36b) \quad \frac{(Q+i\beta\sigma)'_+}{(Q+i\beta\sigma)_+} = -\frac{1}{w-b_2} + \frac{\beta(a_1)}{2\pi(w-a_1)} [G(w) - G(a_1)] - \frac{\beta(a_2)}{2\pi(w-a_2)} \\ \times [G(w) - G(a_2)] + \frac{\beta(b_1)}{2\pi(w-b_1)} [G(w) - G(b_1)] - \frac{\beta(b_2)}{2\pi(w-b_2)} \\ [G(w) - G(b_2)],$$

where

$$F(w) = (1/\beta) [\pi/2 + \sin^{-1} w/\kappa],$$

$$G(w) = (1/\beta) [\pi/2 - \sin^{-1} w/\kappa].$$

Eq. (36) must now be integrated along some contour ending at  $w$ . It will be observed that the integrations are elementary, save for the term involving  $\sin^{-1} w/\kappa$ . This integral may be expressed in terms of the tabulated dilogarithmic integral and related functions.

Noting that

$$P_2(w) = -\kappa^2(w-a_2)(w-b_2) \cos^2 \alpha_2,$$

we find that the final expressions for the factors of  $J(w)$  are

$$(37a) \quad J_-(w) = \exp \left\{ \int_{k_1}^w \left[ \frac{(Q+i\beta\sigma)'_-}{(Q+i\beta\sigma)_-} - \frac{1}{w-a_2} \right] dw \right\},$$

$$(37b) \quad J_+(w) = \frac{1}{J(k_1)} \exp \left\{ \int_{k_1}^w \left[ \frac{(Q+i\beta\sigma)'_+}{(Q+i\beta\sigma)_+} + \frac{1}{w-b_2} \right] dw \right\}.$$

To obtain the factors of  $K$ , we make use of the easily proved relation

$$(38) \quad J(w) K(w) = P_1(w)/P_2(w),$$

where

$$P_1(w) = -\kappa^2(w-a_1)(w-b_1) \cos^2 \alpha_1.$$

Thus,

$$(39a) \quad J_-(w) K_-(w) = c_1(w-a_1)/(w-a_2),$$

$$(39b) \quad J_+(w) K_+(w) = c_2(w-b_2)/(w-b_1),$$

where  $c_1/c_2 = \cos^2 \alpha_1/\cos^2 \alpha_2$ . Since only the ratio  $c_1/c_2$  occurs in the final expressions for the fields, we may define arbitrarily

$$c_1 = \cos \alpha_1/\cos \alpha_2,$$

$$c_2 = \cos \alpha_2/\cos \alpha_1.$$

We conclude this section with a discussion of the asymptotic forms of the factors for large  $|w|$ . The case when either  $\alpha_1$  or  $\alpha_2$  equals  $\pi/2$  is excluded for the time being. Now

$$\sin^{-1} w/\kappa = i \log [(\beta-iw)/\kappa].$$

Thus the dominant term in  $F(w)$  and  $G(w)$  is  $0(\log w/w)$ . Now it can be proved without undue difficulty that

$$(40a) \quad \beta(a_{1,2})F(a_{1,2}) + \beta(b_{1,2})F(b_{1,2}) = \pi - i \sinh^{-1} (2\kappa k_2/\kappa^2),$$

$$(40b) \quad \beta(a_{1,2})G(a_{1,2}) + \beta(b_{1,2})G(b_{1,2}) = \pi + i \sinh^{-1} (2\kappa k_2/\kappa^2).$$

Consequently the coefficient of the  $w^{-1}$  term in the integrands of eqs. (37) vanishes, and we see that the integrands in eqs. (37) are  $0(\log w/w^2)$ .

But

$$\int \log w \cdot dw/w^2 = -(1+\log w)/w \rightarrow 0.$$

Hence  $J_-(w)$  and  $J_+(w)$  tend to constant values as  $|w| \rightarrow \infty$ .

By eqs. (14), (15), and (17a) it can be shown that

$$J(w) \rightarrow c_1, \quad \text{as } |w| \rightarrow \infty,$$

hence if

$$(41a) \quad J_-(w) \rightarrow c, \quad \text{as } |w| \rightarrow \infty$$

then

$$(41b) \quad J_+(w) \rightarrow cc_2, \quad \text{as } |w| \rightarrow \infty;$$

and, using eqs. (39)

$$(42a) \quad K_-(w) \rightarrow c_1/c,$$

$$(42b) \quad K_+(w) \rightarrow 1/c, \quad \text{as } |w| \rightarrow \infty.$$

When  $\alpha_2$  (say) is  $\pi/2$ , the asymptotic forms are quite different. In this case  $a_2, b_2 \rightarrow \infty$ , and  $a_2^{-1} \beta(a_2) + b_2^{-1} \beta(b_2) \rightarrow 0$ , so that

$$(43) \quad \frac{(Q+i\beta\sigma)'}{(Q+i\beta\sigma)_-} \rightarrow -\frac{1}{2\pi} \left\{ \frac{\beta(a_1)}{w-a_1} [F(w) - F(a_1)] + \frac{\beta(b_1)}{w-b_1} [F(w) - F(b_1)] \right\}.$$

Therefore, using eq. (40a)

$$\frac{(Q+i\beta\sigma)'}{(Q+i\beta\sigma)_-} \rightarrow \frac{\frac{1}{2} + \gamma}{w}$$

where

$$\gamma = -(i/2\pi) \sinh^{-1} (2kk_2/\kappa^2).$$

Hence

$$(44a) \quad J_-(w) = O(w^{\frac{1}{2}+\gamma}), \quad \text{for } |w| \rightarrow \infty.$$

In a similar manner

$$(44b) \quad J_+(w) = O(w^{-\frac{1}{2}+\gamma}), \quad \text{for } |w| \rightarrow \infty.$$

#### THE FIELD AT THE BOUNDARY

It is apparent from eqs. (23) that the field quantities are continuous at  $x = 0$  providing  $z \neq 0$ . When  $z = 0$ , however, some tangential field quantities suffer discontinuities across  $x = 0$ . Consider  $I_2(x, 0, p, q)$  and suppose that  $\alpha_1, \alpha_2 \neq \pi/2$ . By addition and subtraction of the asymptotic values of  $J_-(w)$  and  $K_-(w)$ ,  $I_2$  can be expressed in the form

$$(45) \quad I_2(x, 0, p, q) = \int_{-\infty+ia}^{\infty+ia} \frac{e^{iwx}}{w-k_1} \left[ p \left( \frac{1}{J_-(w)} - \frac{1}{c} \right) + q \left( \frac{1}{K_-(w)} - cc_2 \right) \right] dw \\ + \frac{1}{2} \int_{-\infty+ia}^{\infty+ia} \frac{e^{iwx}}{w-k_1} \left[ \frac{p}{c} + qcc_2 \right] dw.$$

The first integrand is  $O(\log w/w^2)$  for large  $|w|$  so that no discontinuity occurs in this term at  $x = 0$ . The second integral can be evaluated by residues, giving

$$(46) \quad \begin{aligned} I &= 2\pi i(pc^{-1} + qcc_2), & x \rightarrow 0_+ \\ &= 0, & x \rightarrow 0_-; \end{aligned}$$

so that by eq. (24b) the discontinuity of  $E_{\xi_2}^{(s)}(x, 0)$  becomes

$$[E_{\xi_2}^{(s)}] \equiv E_{\xi_2}^{(s)}(0_+, 0) - E_{\xi_2}^{(s)}(0_-, 0) = \pi i(pc^{-1} + qcc_2).$$

Similarly,

$$\begin{aligned} [E_{\xi_1}^{(s)}] &= \pi i(pc_1 c^{-1} + qc), \\ [H_{\xi_1}^{(s)}] &= \pi z_0^{-1}(pc_1 c^{-1} - qc), \\ [H_{\xi_2}^{(s)}] &= \pi z_0^{-1}(pc^{-1} - qcc_2). \end{aligned}$$

Hence, using

$$H_x = (-H_{\xi_1} \sin \alpha_2 + H_{\xi_2} \sin \alpha_1) / \sin (\alpha_1 - \alpha_2),$$

$$H_y = (H_{\xi_1} \cos \alpha_2 - H_{\xi_2} \cos \alpha_1) / \sin (\alpha_1 - \alpha_2),$$

we find

$$(47a) \quad [H_x^{(s)}] = \pi z_0^{-1} \sec \alpha_2 (pc^{-1} - qcc_2),$$

$$(47b) \quad [H_y^{(s)}] = 0;$$

and similarly

$$(47c) \quad [E_x^{(s)}] = \pi i \sec \alpha_2 (pc^{-1} + qcc_2),$$

$$(47d) \quad [E_y^{(s)}] = 0.$$

When this problem was first considered, it was anticipated that no unique solution would be found, unless the behavior of the current at the boundary were specified. Considered as the limiting case of an array of wires, two distinct types of current behavior seemed possible, namely, the cases where the conductivity lines do or do not meet at the boundary. The first case corresponds to the current non-zero and continuous in magnitude at the boundary; in the second, the current tends to zero at the boundary. However, it has been shown that only one solution of the problem as formulated is possible, and this solution has the current *discontinuous* in magnitude at the boundary, and thus does not seem to represent either case. One can only doubt the realism of this model to represent the junction of two wire gratings, at least for the fields near the boundary.

#### THE FAR FIELD

To obtain closed-form expressions for the distant field, we have to evaluate integrals of the form

$$(48a) \quad I_1 = \int_{-\infty + ia}^{\infty + ia} \frac{e^{i(wz + \beta z)}}{w - k_1} \left[ \frac{p}{J_+(w)} + \frac{q}{K_+(w)} \right] dw,$$

$$(48b) \quad I_2 = \int_{-\infty+ia}^{\infty+ia} \frac{e^{i(wz+\beta z)}}{w-k_1} \left[ \frac{p}{J_-(w)} + \frac{q}{K_-(w)} \right] dw,$$

when one or both of  $x$  and  $z$  are large.

Consider  $I_2$  and suppose that  $x > 0$ .  $J_-(w)$  and  $K_-(w)$  have branch points at  $w = \kappa$  and zeros at  $w = a_1$  in the upper half-plane. To bring these singularities into evidence, we write

$$\begin{aligned} \frac{1}{J_-(w)} &= \frac{1}{J_+(w)J(w)} = \frac{K(w)P_2(w)}{J_+(w)P_1(w)}, & (\text{by eq. (38)}), \\ &= -\frac{Q(w)-i\beta\sigma}{\kappa^2(w-a_1)(w-b_1)J_+(w)\cos^2\alpha_1}. \end{aligned}$$

Similarly,

$$\frac{1}{K_-(w)} = -\frac{Q(w)+i\beta\sigma}{\kappa^2(w-a_1)(w-b_1)K_+(w)\cos^2\alpha_1}$$

so that

$$(49) \quad I_2 = -\frac{1}{\kappa^2\cos^2\alpha_1} \int_{-\infty+ia}^{\infty+ia} \frac{e^{i(wz+\beta z)}}{(w-k_1)(w-a_1)(w-b_1)} \left\{ Q(w) \left[ \frac{p}{J_+(w)} + \frac{q}{K_+(w)} \right] - i\beta\sigma \left[ \frac{p}{J_+(w)} - \frac{q}{K_+(w)} \right] \right\} dw.$$

Put

$$(50a) \quad U(w) = -\frac{(\kappa+w)^{\frac{1}{2}}Q(w)}{\kappa^2(w-b_1)\cos^2\alpha_1} \left[ \frac{p}{J_+(w)} + \frac{q}{K_+(w)} \right],$$

$$(50b) \quad V(w) = \frac{i\sigma\beta(\kappa-w)^{\frac{1}{2}}}{\kappa^2(w-b_1)\cos^2\alpha_1} \left[ \frac{p}{J_+(w)} - \frac{q}{K_+(w)} \right].$$

These functions are regular in the upper half-plane. Then

$$I_2 = \int_{-\infty+ia}^{\infty+ia} \frac{e^{i(wz+\beta z)}}{k_1-a_1} \left[ \frac{U(w)}{(\kappa+w)^{\frac{1}{2}}} + \frac{V(w)}{(\kappa-w)^{\frac{1}{2}}} \right] \left[ \frac{1}{w-k_1} - \frac{1}{w-a_1} \right] dw,$$

which can be rewritten as

$$\begin{aligned} (51) \quad I_2 &= \int_{-\infty+ia}^{\infty+ia} \frac{e^{i(wz+\beta z)}}{k_1-a_1} \left[ \frac{U(w)-U(k_1)}{(w-k_1)(\kappa+w)^{\frac{1}{2}}} + \frac{V(w)-V(k_1)}{(w-k_1)(\kappa-w)^{\frac{1}{2}}} \right] dw \\ &\quad - \int_{-\infty+ia}^{\infty+ia} \frac{e^{i(wz+\beta z)}}{k_1-a_1} \left[ \frac{U(w)-U(a_1)}{(w-a_1)(\kappa+w)^{\frac{1}{2}}} + \frac{V(w)-V(a_1)}{(w-a_1)(\kappa-w)^{\frac{1}{2}}} \right] dw \\ &\quad + \int_{-\infty+ia}^{\infty+ia} \frac{e^{i(wz+\beta z)}}{k_1-a_1} \left[ \frac{U(k_1)(\kappa+w)^{-\frac{1}{2}} + V(k_1)(\kappa-w)^{-\frac{1}{2}}}{w-k_1} \right. \\ &\quad \left. - \frac{U(a_1)(\kappa+w)^{-\frac{1}{2}} + V(a_1)(\kappa-w)^{-\frac{1}{2}}}{w-a_1} \right] dw. \end{aligned}$$

The first two integrals in eq. (51) can be evaluated by stationary phase since their integrands are free of poles in  $v > -q$ ; the last integral can be evaluated exactly in terms of Fresnel integrals (see Hurd 1960).

The result is

$$(52) \quad I_2 \simeq 2 \left( \frac{\pi}{\rho} \right)^{\frac{1}{2}} \frac{e^{i(\kappa\rho - \pi/4)}}{\kappa(k_1 - a_1)} \left\{ \frac{[U(\kappa \cos \phi) - U(\kappa \cos \theta)] \sin \frac{1}{2}\phi + [V(\kappa \cos \phi) - V(\kappa \cos \theta)] \cos \frac{1}{2}\phi}{\cos \phi - \cos \theta} \right. \\ - \frac{[U(\kappa \cos \phi) - U(\kappa \cosh \psi_1)] \sin \frac{1}{2}\phi + [V(\kappa \cos \phi) - V(\kappa \cosh \psi_1)] \cos \frac{1}{2}\phi}{\cos \phi - \cosh \psi_1} \Big\} \\ + \left( \frac{2\pi}{\kappa} \right)^{\frac{1}{2}} \frac{e^{i\pi/4}}{k_1 - a_1} \left\{ \frac{U(k_1)[H(\phi, \theta) + H(\phi, -\theta)]}{\cos \frac{1}{2}\theta} - \frac{V(k_1)[H(\phi, \theta) - H(\phi, -\theta)]}{\sin \frac{1}{2}\theta} \right. \\ \left. - \frac{U(a_1)[H(\phi, i\psi_1) + H(\phi, -i\psi_1)]}{\cosh \frac{1}{2}\psi_1} - \frac{iV(a_1)[H(\phi, i\psi_1) - H(\phi, -i\psi_1)]}{\sinh \frac{1}{2}\psi_1} \right\},$$

where

$$\rho^2 = x^2 + z^2,$$

$$\phi = \cos^{-1} x/\rho,$$

$$\theta = \cos^{-1} k_1/\kappa,$$

$$\psi_1 = \cosh^{-1} a_1/\kappa,$$

and

$$H(\phi, \theta) = e^{i\kappa\rho \cos(\phi+\theta)} \int_{-\infty}^{-\sqrt{2\kappa\rho} \sin \frac{1}{2}(\phi+\theta)} e^{i\lambda^2 d\lambda}.$$

The situation is much simpler when  $x < 0$ , since  $J_-(w)$  and  $K_-(w)$  are regular in the lower half  $w$ -plane.

Put

$$W(w) = (\kappa - w)^{\frac{1}{2}} [pJ_-^{-1}(w) + qK_-^{-1}(w)],$$

so that

$$I_2 = \int_{-\infty + ia}^{\infty + ia} \frac{e^{i(wx + \beta z)} [W(w) - W(k_1)]}{(\kappa - w)^{\frac{1}{2}} (w - k_1)} dw + W(k_1) \int_{-\infty + ia}^{\infty + ia} \frac{e^{i(wx + \beta z)}}{(\kappa - w)^{\frac{1}{2}} (w - k_1)} dw.$$

Therefore

$$(53) \quad I_2 \simeq \frac{2}{\kappa} \left( \frac{\pi}{\rho} \right)^{\frac{1}{2}} e^{i(\kappa\rho - \pi/4)} \cos \frac{1}{2}\phi \left[ \frac{W(\kappa \cos \phi) - W(\kappa \cos \theta)}{\cos \phi - \cos \theta} \right. \\ \left. - W(k_1) \left( \frac{2\pi}{\kappa} \right)^{\frac{1}{2}} \frac{e^{i\pi/4}}{\sin \frac{1}{2}\theta} [H(\phi, \theta) - H(\phi, -\theta)] \right].$$

The analysis follows along similar lines for  $I_1$ , with the result for  $x > 0$ ,

$$(54) \quad I_1 \simeq \frac{2}{\kappa} \left( \frac{\pi}{\rho} \right)^{\frac{1}{2}} e^{i(\kappa\rho - \pi/4)} \sin \frac{1}{2}\phi \left[ \frac{X(\kappa \cos \phi) - X(\kappa \cos \theta)}{\cos \phi - \cos \theta} \right. \\ \left. + X(k_1) \left( \frac{2\pi}{\kappa} \right)^{\frac{1}{2}} \frac{e^{i\pi/4}}{\cos \frac{1}{2}\theta} [H(\phi, \theta) + H(\phi, -\theta)]; \right]$$



while for  $x < 0$ ,

$$(55) \quad I_1 = 2 \left( \frac{\pi}{\rho} \right)^{\frac{1}{2}} \frac{e^{i(\kappa\rho - \pi/4)}}{\kappa(k_1 - b_2)} \left\{ \frac{[Y(\kappa \cos \phi) - Y(\kappa \cos \theta)] \cos \frac{1}{2} \phi + [Z(\kappa \cos \phi) - Z(\kappa \cos \theta)] \sin \frac{1}{2} \phi}{\cos \phi - \cos \theta} \right. \\ + \frac{[Y(\kappa \cos \phi) - Y(\kappa \cosh \psi_2)] \cos \frac{1}{2} \phi + [Z(\kappa \cos \phi) - Z(\kappa \cosh \psi_2)] \sin \frac{1}{2} \phi}{\cos \phi - \cosh \psi_2} \Big\} \\ + \left( \frac{2\pi}{\kappa} \right)^{\frac{1}{2}} \frac{e^{i\pi/4}}{k_1 - b_2} \left\{ \frac{Y(k_1)[H(\phi, \theta) + H(\phi, -\theta)]}{\cos \frac{1}{2} \theta} - \frac{Y(k_1)[H(\phi, \theta) - H(\phi, -\theta)]}{\sin \frac{1}{2} \theta} \right. \\ \left. - \frac{Z(b_2)[H(\phi, i\psi_2) + H(\phi, -i\psi_2)]}{\cosh \frac{1}{2} \psi_2} - \frac{iY(b_2)[H(\phi, i\psi_2) - H(\phi, -i\psi_2)]}{\sinh \frac{1}{2} \psi_2} \right\}$$

where

$$(56a) \quad X(w) = (\kappa + w)^{\frac{1}{2}} [pJ_+^{-1}(w) + qK_+^{-1}(w)],$$

$$(56b) \quad Y(w) = -\frac{(\kappa - w)^{\frac{1}{2}} Q(w)}{\kappa^2 \cos^2 \alpha_2 (w - a_2)} \left[ \frac{p}{J_-(w)} + \frac{q}{K_-(w)} \right],$$

$$(56c) \quad Z(w) = \frac{-i\beta\sigma(\kappa + w)^{\frac{1}{2}}}{\kappa^2 \cos^2 \alpha_2 (w - a_2)} \left[ \frac{p}{J_-(w)} - \frac{q}{K_-(w)} \right],$$

and

$$\psi_2 = \cosh^{-1} b_2/\kappa.$$

We note that the far field formulae have terms of three types. First, there are the ordinary diffracted waves decaying as  $\rho^{-\frac{1}{2}}$ ; then there are the usual shadow-forming waves which are described by  $H(\phi, \theta)$  and  $H(\phi, -\theta)$ . Finally there are surface wave terms contained in  $H(\phi, i\psi)$ ,  $H(\phi, -i\psi)$ . The surface waves propagate according to  $e^{i(a_1 x + \beta(a_1)z)}$  in  $x > 0$  and  $e^{i(b_2 x + \beta(b_2)z)}$  in  $x < 0$ .

#### BABINET'S PRINCIPLE

It is a consequence of the work of Toraldo di Francia (1956b) that the surface which we are considering is its own Babinet complement; that is, if there are two incident fields  $\mathbf{E}^{1,1}$  and  $\mathbf{E}^{2,1}$  given by

$$(58a) \quad \mathbf{E}^{1,1} = \mathbf{A}e^{i\mathbf{k}\cdot\mathbf{r}}, \quad z_0 \mathbf{H}^{1,1} = k^{-1}(\mathbf{k} \times \mathbf{A})e^{i\mathbf{k}\cdot\mathbf{r}},$$

$$(58b) \quad \mathbf{E}^{2,1} = -z_0 \mathbf{H}^{1,1}, \quad z_0 \mathbf{H}^{2,1} = \mathbf{E}^{1,1};$$

then the relations

$$\mathbf{E}^1 + z_0 \mathbf{H}^2 = \mathbf{E}^{1,1},$$

$$z_0 \mathbf{H}^1 - \mathbf{E}^2 = \mathbf{H}^{1,1}$$

must hold. Or equivalently,

$$(59a) \quad \Pi_y^1 + z_0 \Pi_y^{2*} = \Pi_y^{1,1},$$

$$(59b) \quad z_0 \Pi_y^{1*} - \Pi_y^2 = z_0 \Pi_y^{1,1*}.$$

We have verified that the solution (eqs. (23)) satisfies eqs. (59), but since the algebra is a bit tedious, it is not included here.

## A PARTICULAR CASE

There is only one special case for which the expressions show any sizable simplification. This is when  $\alpha_1 = \alpha_2 = \alpha$ . In this case, from eqs. (14) and (15)

$$P_2(w) = Q(w).$$

Also,  $\sigma = 0$ , so that  $J(w) = K(w) = 1$ .

Therefore, by eqs. (39), (40),

$$J_-(w) = J_+(w) = c,$$

$$K_-(w) = K_+(w) = c^{-1}.$$

Thus  $p = q = 0$ , and therefore  $I_1 = I_2 = 0$ . We are left with

$$(60a) \quad \Pi_y^{(s)} = (k_1 k_2 \cos \alpha - \kappa^2 \sin \alpha) P_2^{-1}(k_1) A_\xi e^{i(k_1 x + k_3 z)},$$

$$(60b) \quad \Pi_y^{(s)*} = -k k_3 \varepsilon_0^{-1} P_2^{-1}(k_1) A_\xi \cos \alpha \cdot e^{i(k_1 x + k_3 z)},$$

and it is easy to show that the fields defined by eqs. (60) satisfy the boundary conditions over the whole plane.

## ACKNOWLEDGMENT

The author is indebted to Dr. R. F. Millar for reading the manuscript and for many helpful discussions of the problem.

## REFERENCES

- BOUWKAMP, C. J. 1954. Repts. Progr. in Phys. **17**, 35.  
 HURD, R. A. 1960. Can. J. Phys. **38**, 168.  
 KARP, S. N. 1957. New York Institute of Mathematical Sciences, Report EM-108.  
 RADLOW, J. 1959. Quart. Appl. Math. **17**, 113.  
 TORALDO DI FRANCIA, G. 1956a. Nuovo Cimento, **3**, 1276.  
 ——— 1956b. R.C. Accad. Naz. Lincei, **20**, 476.

# CONSIDERATIONS ON A MESON-ATOMIC MODEL OF THE NUCLEON<sup>1</sup>

L. E. H. TRAINOR

## ABSTRACT

A model of the nucleon is described in which a  $\pi$ -meson moves about a nucleon core under the action of a hyper-Maxwell field. On such a model, the short range of the internucleon force appears as a screening effect. Despite its obvious limitations, the model does possess some interesting features which lead to results in agreement with experiment. The advantage to such models is that they may provide insight into problems which are enormously difficult from the usual field theory point of view. In particular, one might hope to obtain some understanding of the state of the  $\pi$ -meson field in the nuclear many-body problem.

## I. INTRODUCTION

Feld (1957, 1958) has shown that it is possible to construct atomic models for the physical nucleons which reproduce the static properties of the neutron, proton, and deuteron, and which provide a rough description of such processes as the photoproduction of mesons and the scattering of  $\pi$ -mesons by nucleons. Feld suggests that to the extent that such a model is successful it may be regarded as a useful approximation to the complete field theory description. A similar point of view is adopted here. Feld makes the *ad hoc* assumption that a single  $\pi$ -meson moves in a  $p$ -state about a spin  $1/2$  core, the total angular momentum being  $1/2$ . The forces between core and meson are not specified in detail, nor is any account taken of the  $\pi$ - $\pi$  interaction. Unfortunately, not very much is known about the  $\pi$ - $\pi$  interaction, but it is expected to display a strong  $p$ -wave resonance in a state of isotopic spin  $T = 1$  (Chew 1959).

In the present work a more explicit model than Feld's is proposed (Trainor 1959), a model which is certainly not correct, but which has certain features of considerable interest: It provides a ground state which is  $p_{1/2}$  with total energy zero, a feature which is essential to an atomic model description of  $\pi$ -meson production. It also provides in a unified description a strong  $p$ -wave interaction between  $\pi$ -mesons on the one hand, and between  $\pi$ -mesons and nucleons on the other. Finally, it suggests the possibility that in the final analysis, all basic forces in nature may be long range forces transmitted by fields of massless quanta.

The model is introduced in Section II and the solutions of interest are derived in Section III. The physical admissibility of these solutions is examined in Section IV. In Section V, applications of the model are briefly discussed. In particular, consideration is given to the free-meson lattice model of nuclear matter (Trainor and Sood 1959), in which  $\pi$ -mesons become dissociated from individual nucleons and move through a quasi lattice of nucleon cores, some-

<sup>1</sup>Manuscript received May 6, 1960.

Contribution from the Department of Physics, University of Alberta, Edmonton, Alberta.

what in the manner of conduction electrons in a metal lattice. It is at once evident that on such a model, the properties of nuclear matter involve the  $\pi$ - $\pi$  interaction as well as the  $\pi$ -nucleon interaction in an essential way.

## II. AN ATOMIC MODEL

Consider, for simplicity, that  $\pi$ -mesons and nucleon cores interact through the medium of a hyper-Maxwell field. Let nucleon cores be "hypercharged" with a charge or strength of  $g$ , and  $\pi$ -mesons with a strength of  $-g$ . In the simplest view, the physical nucleon consists of a single  $\pi$ -meson circulating around a nucleon core under the action of the attractive hyper-Maxwell field. If one were to interpret the situation literally, there would exist in nature a field particle, the hyperphoton, of zero mass and spin 1 obeying a set of Maxwell-like equations.

Case (1950) and Feshbach and Villars (1958) have shown that the single-component Klein-Gordon equation, regarded from the single particle point of view, corresponds to a particle having two internal degrees of freedom. In the presence of a weak Maxwell field, the two degrees of freedom are distinguished as two possible states of the charge  $\pm e$ . In a strong Maxwell field, the single particle picture is lost in the appearance or creation of particle pairs having opposite signs of the charge  $e$ .

The Feshbach and Villars representation is adopted here for describing the motion of the  $\pi$ -meson in the hyper-Maxwell field of the nucleon core. To avoid the hopeless complication of the many-body problem, let us assume a single  $\pi$ -meson and ignore the  $K$ -meson structure of the nucleon core. Further, to avoid the equally hopeless complication of the relativistic two-body problem, let us solve the problem in the approximation that the core motion is neglected. In other words, our problem is just the Kepler problem for a scalar (or pseudoscalar)  $\pi$ -meson of hypercharge  $-g$  moving about an attractive center of hypercharge  $g$  under the action of a static potential  $\phi = -g^2/r$ . This problem has been solved many times, particularly in the limit of small  $g$  (see, for example, Schiff (1955)), but also for large  $g$  values (Case 1950). It is proposed here that for certain  $g$  values corresponding to real eigenvalues of the energy, there are solutions to the Klein-Gordon equation that are usually discarded, which may be of considerable physical interest.

In considering possible atomic models for the physical nucleon, it is important to obtain a ground state corresponding to a real meson circulating in a state of zero total energy, since the experimentally observed threshold for meson photoproduction corresponds to the  $\pi$ -meson rest mass. It will be shown that the Klein-Gordon equation does possess unusual solutions having this property for a few restricted choices of the interaction strength  $g^2/\hbar c$ . For the particular value  $\sqrt{2}$  for this quantity, the lowest bound state of the  $\pi$ -meson corresponds to a  $\pi$ -meson of total energy zero moving in a  $p$ -orbit, the system as a whole having angular momentum  $1/2$ . This result enables one to account for the static properties of the physical nucleons in a reasonably satisfactory way (Feld 1957, 1958).

On our model, photoproduction corresponds to the ordinary photoelectric ejection of a  $\pi$ -meson from a bound state of total energy zero to a free state of energy  $mc^2$ , where  $m$  is the dynamical mass. The subsequent behavior of the stripped core requires separate discussion. For simplicity one can imagine that in the over-all process, hypercharge is not conserved so that the stripped core simply picks up a meson from the vacuum and becomes once more a physical nucleon, corresponding to a zero energy  $\pi$ -meson circulating around a nucleon core. The whole event then amounts to the process of gamma rays pumping  $\pi$ -mesons out of the vacuum in the presence of a physical nucleon.

### III. UNUSUAL SOLUTIONS OF THE KLEIN-GORDON KEPLER PROBLEM

The Klein-Gordon equation is given by (Feshbach and Villars 1958)

$$(1) \quad \sum_{\mu=1}^4 \left( p_{\mu} - \frac{g}{c} A_{\mu} \right)^2 \Psi + m^2 c^4 \Psi = \sum_{\mu=1}^4 (D_{\mu}^2 - \kappa^2) \Psi = 0$$

where

$$D_{\mu} = \frac{\partial}{\partial x_{\mu}} - \frac{ig}{\hbar c} A_{\mu}, \quad \kappa = \frac{mc}{\hbar}, \quad \text{and} \quad p_4 = -\frac{\hbar}{c} \frac{\partial}{\partial t},$$

and where the current

$$(2) \quad J_{\mu} = [\Psi^* D_{\mu} \Psi - (D_{\mu}^* \Psi^*) \Psi]$$

satisfies the continuity equation

$$(3) \quad \sum_{\mu=1}^4 \frac{\partial J_{\mu}}{\partial x_{\mu}} = 0.$$

The Kepler problem is defined by the choice

$$(4) \quad A_{\mu} = 0, \mu = 1, 2, 3; \quad A_4 = i\phi = -ig/r.$$

To find the states of definite energy, one separates off the time dependence in the usual way to obtain the Klein-Gordon equation independent of the time (see Schiff 1955):

$$(5) \quad \Psi(\mathbf{r}, t) = \psi(\mathbf{r}) e^{-i\epsilon t/\hbar},$$

$$(6) \quad (-\hbar^2 c^2 \nabla^2 + m^2 c^4) \psi(\mathbf{r}) = (\epsilon - g\phi)^2 \psi(\mathbf{r}).$$

This equation can be separated in spherical co-ordinates

$$(7) \quad r \psi_{lq}(\mathbf{r}) = u_l(r) Y_{lq}(\theta, \varphi).$$

The radial equation becomes

$$(8) \quad \frac{d^2 u_l}{dr^2} + \left[ \frac{\gamma^2 - l(l+1)}{r^2} + \frac{\lambda \gamma}{r} - \alpha^2 \right] u_l = 0$$

where we have put

$$(9) \quad \phi = -g/r, \quad \lambda = 2\epsilon/\hbar c, \quad \alpha^2 = \kappa^2 - \lambda^2/4, \quad \text{and} \quad \gamma = g^2/\hbar c.$$

After a simple transformation, equation (8) becomes the confluent hypergeometric equation of Whittaker whose solutions can be written down at once. However, for the solutions of interest to us, it is simpler to start from first principles.

As  $r \rightarrow \infty$ ,  $u_l(r) \rightarrow e^{-\alpha r}$  with  $\alpha$  positive for bound states. Substituting  $u_l = f_l(r)e^{-\alpha r}$  yields

$$(10) \quad \frac{d^2 f_l}{dr^2} - 2\alpha \frac{df_l}{dr} + \frac{(\gamma^2 - l(l+1))}{r^2} f_l + \frac{\gamma\lambda}{r} f_l = 0.$$

For a power series solution

$$(11) \quad f_l = \sum_{n=0}^{\infty} a_n r^{n+s},$$

the indicial equation yields an exponent

$$(12) \quad s = \frac{1}{2} \pm \sqrt{(l + \frac{1}{2})^2 - \gamma^2}.$$

The recurrence formula is

$$(13) \quad a_{n+1} = \left( \frac{2\alpha(n+s) - \gamma\lambda}{(n+s+1)(n+s) + \gamma^2 - l(l+1)} \right) a_n.$$

This series diverges as  $e^{2\alpha r}$  for large  $r$  so that only the polynomial solutions are acceptable. To cut off the series at the  $j$ th term requires  $a_j \neq 0$ ,  $a_{j+1} = 0$ , hence,

$$(14) \quad \gamma\lambda_j = 2\alpha_j(j+s).$$

From equations (9) and (14) it follows that

$$\lambda_j = 2(j+s)\kappa/(\gamma^2 + (j+s)^2)^{\frac{1}{2}}.$$

Hence

$$(15) \quad \epsilon_j = (j+s)mc^2/(\gamma^2 + (j+s)^2)^{\frac{1}{2}} \quad \text{and} \quad \alpha_j = \gamma\kappa/(\gamma^2 + (j+s)^2)^{\frac{1}{2}}.$$

From the indicial equation we see an essential difficulty with the Klein-Gordon equation. If the interaction becomes sufficiently strong,  $s$  becomes complex for certain values of  $l$ . It then follows from equation (15) that the energy eigenvalues are complex. This conclusion is justified by a more thorough examination of the general solutions of the confluent hypergeometric equation. In other words, these solutions correspond to states which do not conserve energy and they reflect the fact that our Hamiltonian is not Hermitian in the usual sense. From the point of view of field theory, this represents the fact that the interaction is strong enough to destroy the single particle picture through a vacuum polarization near  $r = 0$ . Due to the repulsive nature of the centripetal potential, the breakdown of the single particle picture is less serious for states of higher angular momenta. Here we adopt a strict single particle point of view and rule out states with complex energy eigenvalues as physically inadmissible. The requirement that  $s$  be real leads to

$$(16) \quad \gamma^2 \leq (l + \frac{1}{2})^2$$

for either choice of sign in equation (12).

We also require  $s > 0$  for square integrability of the wave functions. If the upper sign is chosen for  $s$ , then  $s \geq \frac{1}{2}$ ; if the lower sign is chosen we are led to the condition

$$(17) \quad (l + \frac{1}{2})^2 - \gamma^2 < \frac{1}{4}.$$

The two conditions (16) and (17) on  $s$  imply restrictions on the value of  $\gamma = g^2/\hbar c$  for any particular orbital momentum  $l$ . These restrictions are summarized in Table I where the two sets of solutions, corresponding to the

TABLE I

Range of permissible values of the interaction strength  $\gamma^2$  for several values of the orbital angular momentum quantum number  $l$ . These ranges are established by the requirements that  $s$  be both positive and real.  $s^{(1)}$  corresponds to the usual solutions,  $s^{(2)}$  to the unusual solutions

Orbital momentum, $l$	Range of permissible values of the interaction strength $\gamma^2$	
	Branch $s^{(1)}$	Branch $s^{(2)}$
0	0- 0.25	0 - 0.25
1	0- 2.25	2.00- 2.25
2	0- 6.25	6.00- 6.25
3	0-12.25	12.00-12.25
4	0-20.25	20.00-20.25

two choices of sign in equation (12), are distinguished as belonging to branch  $s^{(1)}$  or  $s^{(2)}$ . One notes that if  $s$ -states are to occur  $\gamma^2$  must be less than 0.25, while if  $p$ -states are to occur,  $\gamma^2$  must be less than 2.25.

Table II and Fig. 1 give the variation of the energy eigenvalues with interaction strength  $\gamma^2$  for both branches of the indicial equation, for angular

TABLE II

Variation of energy eigenvalues with interaction strength. Various energy levels are denoted by  $\epsilon_{j,l}^{(\sigma)}$  where  $l$  gives the angular momentum,  $j$  the principal quantum number, and  $\sigma = 1, 2$  denotes the usual and unusual branches respectively. Energies are measured in units of  $mc^2$

Interaction strength, $\gamma^2$	Energy eigenvalues							
	$\epsilon_{00}^{(1)}$	$\epsilon_{00}^{(2)}$	$\epsilon_{10}^{(1)}$	$\epsilon_{10}^{(2)}$	$\epsilon_{01}^{(1)}$	$\epsilon_{01}^{(2)}$	$\epsilon_{11}^{(1)}$	$\epsilon_{11}^{(2)}$
0.01	0.995	0.101	0.999	0.996	0.999		0.999	
0.10	0.942	0.336	0.986	0.962	0.987		0.993	
0.15	0.903	0.429	0.978	0.950	0.980		0.989	
0.20	0.851	0.526	0.969	0.944	0.976		0.987	
0.22	0.820	0.572	0.963	0.943	0.973		0.986	
0.25	0.707	0.707	0.949	0.949	0.968		0.986	
1.00					0.851		0.934	
1.50					0.745		0.888	
2.01					0.572	0.007	0.814	0.580
2.10					0.522	0.078	0.793	0.609
2.15					0.486	0.124	0.778	0.628
2.20					0.444	0.183	0.758	0.652
2.22					0.412	0.214	0.747	0.665
2.25					0.316	0.316	0.707	0.707

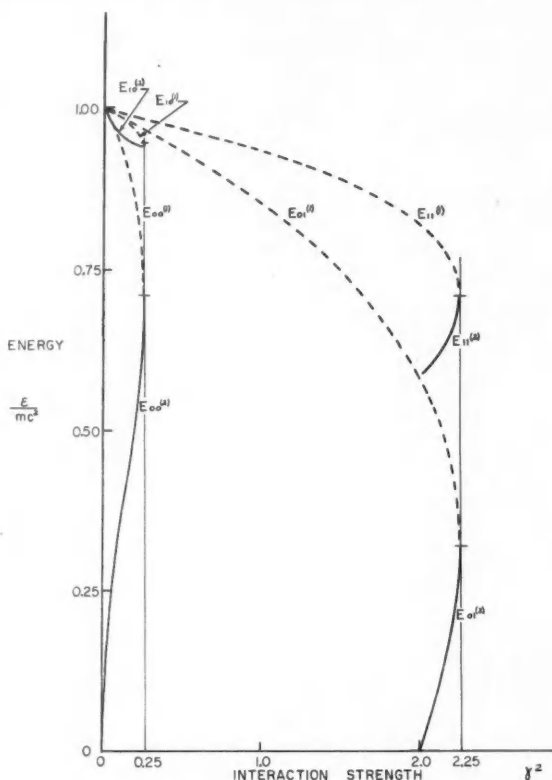


FIG. 1. Variation of energy levels with interaction strength  $\gamma^2$  for a Klein-Gordon particle moving in an inverse square field of force. Energy levels are denoted by  $\epsilon_{j,l}^{(\sigma)}$  as described in the text.

momenta  $l = 0$  and  $l = 1$  and for "principal" quantum numbers  $j = 0$  and  $1$ . The notation  $\epsilon_{j,l}^{(\sigma)}$  is used, where  $\sigma = 1, 2$  denotes the usual and unusual branches, respectively, corresponding to the choice of positive or negative sign in equation (12) for  $s$ . Table III records the values of  $\alpha_{j,l}^{(\sigma)}$  for the same states. This is the parameter which determines the exponential decay of the wave functions for large values of  $r$ .

Of particular interest here is the fact that as  $\gamma^2$  approaches  $2$  from above, the unusual branch solution corresponding to  $j = 0, l = 1$ , has an energy eigenvalue approaching zero. For this value of  $\gamma^2$  there are no  $s$ -states, at least in the strict single particle picture, so that this  $p$ -state of zero total energy represents the ground state of the system. In the present context, this state is to be regarded as describing the motion of the  $\pi$ -meson in the physical nucleon. The apparent parity change usually associated with the creation of a



TABLE III

Variation of the decay constants in units of  $mc/\hbar$  with interaction strength  $\gamma^2$ . Asymptotically the wave functions of the energy eigenstates fall off as  $\exp(-\alpha_{j,l}^{(\sigma)}r)$ . Here  $l$  is the angular momentum of the state,  $j$  is its principal quantum number, and  $\sigma = 1, 2$  designates the branch of solutions

Interaction strength, $\gamma^2$	Asymptotic decay constants							
	$\alpha_{00}^{(1)}$	$\alpha_{00}^{(2)}$	$\alpha_{10}^{(1)}$	$\alpha_{10}^{(2)}$	$\alpha_{01}^{(1)}$	$\alpha_{01}^{(2)}$	$\alpha_{11}^{(1)}$	$\alpha_{11}^{(2)}$
0.01	0.101	0.995	0.025	0.099				
0.10	0.336	0.942	0.089	0.273				
0.15	0.429	0.903	0.209	0.311				
0.20	0.526	0.851	0.251	0.331				
0.22	0.572	0.820	0.270	0.333				
0.25	0.707	0.707	0.316	0.316				
2.01					0.820	1.000	0.580	0.814
2.10					0.853	0.997	0.609	0.793
2.15					0.874	0.992	0.628	0.778
2.20					0.899	0.983	0.652	0.758
2.22					0.913	0.979	0.667	0.748
2.25					0.949	0.949	0.707	0.707

pseudoscalar meson appears here as the natural consequence of the  $l = 1$  ground state of the meson (scalar or pseudoscalar) in the atomic model.

#### IV. ADMISSIBILITY OF SOLUTIONS

It is necessary to examine with some care the admissibility of these unusual solutions belonging to the second branch of the indicial equation. A discussion of the boundary conditions which must be satisfied by the solutions of a Schrodinger equation can be found in many places. (See, for example, Landau and Lifshitz (1958).) The usual (non-relativistic) requirements are: the wave functions must be finite and continuous at all points in space, they must be square integrable, and they must satisfy the usual orthogonality conditions. An exception to the requirement of finiteness is necessary if the potential has a singularity at any point. There is then no physical argument to prevent the probability amplitude from going to infinity at the point in question, provided that square integrability still holds, and that orthogonality is maintained. This latter requirement is equivalent, as will be seen shortly, to the requirement that the radial probability current in the neighborhood of the singularity remains finite.

Following Feshbach and Villars (1958) one may introduce a two-component wave function,

$$(18) \quad \Phi = \begin{pmatrix} \varphi \\ \chi \end{pmatrix}$$

where

$$(19) \quad \begin{aligned} \varphi &= \frac{1}{\sqrt{2}mc^2} \left( mc^2 + \epsilon + \frac{g^2}{r} \right) \Psi, \\ \chi &= \frac{1}{\sqrt{2}mc^2} \left( mc^2 - \epsilon - \frac{g^2}{r} \right) \Psi. \end{aligned}$$

Here  $\Psi$  satisfies the Klein-Gordon equation (1). In this representation, the Hamiltonian

$$(20) \quad H = \frac{p^2}{2m} (\tau_3 + i\tau_2) + \tau_3 mc^2 - \frac{g^2}{r},$$

where the  $\tau$  are the usual  $2 \times 2$  matrices of the Pauli type, leads to a Schrodinger equation linear in the time,

$$(21) \quad H\Phi = i\hbar(\partial\Phi/\partial t).$$

The two degrees of freedom correspond to the two signs of the "hypercharge". The normalization and orthogonality relations between eigenfunctions would be expected to have the form

$$(22) \quad \int \Phi_1^* \tau_3 \Phi_2 d^3r = \pm \delta_{12},$$

which may be expressed in terms of the one-component functions  $\varphi$  and  $\chi$

$$(23) \quad \int (\varphi_2^* \varphi_1 - \chi_2^* \chi_1) d^3r = \pm \delta_{12}.$$

From the differential equations (Feshbach and Villars 1958) satisfied by  $\varphi$  and  $\chi$ , one obtains

$$(24) \quad (\epsilon_1 - \epsilon_2) \int (\varphi_2^* \varphi_1 - \chi_2^* \chi_1) d^3r = \pm \delta_{12} = -\frac{\hbar^2}{2m} \int (\psi_2^* \nabla^2 \psi_1 - \psi_1 \nabla^2 \psi_2^*) d^3r.$$

Here  $\psi_1, \psi_2$  are time-independent solutions of the Klein-Gordon equation corresponding to energy eigenvalues  $\epsilon_1$  and  $\epsilon_2$ . Integrating the right-hand side of this equation over angles one obtains for the orthogonality condition,

$$(25) \quad \int_0^\infty (u_2^* u_1'' - u_2''^* u_1) dr = 0$$

where  $\psi_1 = ru_1, \psi_2 = ru_2$ , and primes denote differentiation with respect to  $r$ . Hence,

$$(26) \quad [u_2^* u_1' - u_2'^* u_1]_0^\infty = 0.$$

Since the solutions all decay exponentially at infinity, one has the condition

$$(27) \quad \lim_{r \rightarrow 0} (u_2^* u_1' - u_2'^* u_1) = 0.$$

Let solution  $u_2$  behave as  $r^{s_2} e^{-\alpha_2 r}$  and solution  $u_1$  behave as  $r^{s_1} e^{-\alpha_1 r}$  near  $r = 0$ . It then follows that if  $s_1 \neq s_2$  (i.e. if the two solutions belong to different branches of the indicial equation) the orthogonality condition is not satisfied. If, however,  $s_1 = s_2 = s$ , so that both solutions belong to the same branch of the indicial equation, orthogonality follows. The orthogonality of states belonging to different values of the angular momentum quantum numbers is ensured by the angular integrations, without regard to which branch they belong.

One has then the situation that either branch of the indicial equation leads to solutions which behave satisfactorily at  $r = 0$  and satisfy the orthogonality

relations. However, for a given angular momentum the solutions of one branch are not orthogonal to those of the other branch. (If  $\gamma = \sqrt{2}$  only the  $p$ -states are affected, since only  $l = 1$  solutions then occur in the unusual branch.) This lack of orthogonality between states with  $l = 1$  destroys the usual statistical interpretation associated with a linear superposition of orthogonal states in quantum mechanics, but this interpretation is not essential to our applications.

It should be noted that one cannot circumvent the problem of orthogonality by ruling out the unusual branch solutions. In fact, these solutions are physically as well-behaved as the usual branch solutions, and there is no obvious way of discriminating against either without discriminating against both.

#### V. APPLICATIONS

Costa and Feld (1960) have already demonstrated that an atomic model in which a  $\pi$ -meson moves around a nucleon core in a  $p_{1/2}$  state is capable of yielding certain gross features of the cross sections for photo-meson production and for  $\pi$ -nucleon scattering because of the dominance of the (3,3) resonance. Their calculation of the scattering of  $\pi$ -mesons on nucleons, however, does not take into account the  $\pi$ - $\pi$  interaction. This calculation is also being carried out using the model suggested in Section II of the present paper. It will be interesting to see what effects arise from the inclusion of a  $\pi$ - $\pi$  interaction which has the same strength as the  $\pi$ -core interaction.

In connection with  $\pi$ -nucleon scattering, it should be noted that the circumstances described in Section IV regarding the bound states of our atomic model hold also for the positive energy solutions. For the choice of interaction strength  $\gamma = \sqrt{2}$ , no  $s$ -states occur which conserve energy. For  $p$ -states both branches of the indicial equation contribute, while for higher angular momenta only a single branch contributes. It is evident from this that the scattering at moderate energies will be dominated to a very great extent by  $p$ -waves, in accord with experiment.

Another application of the atomic model of the nucleon described above is to provide some insight into the free-meson lattice model of nuclear matter (Trainor and Sood 1959). According to this model,  $\pi$ -mesons in nuclear matter become dissociated from individual nucleons, and move through a quasi lattice of nucleon cores in a manner reminiscent of the motion of free electrons in a metal.

Our approach has been to use the atomic model of Sections II and III in a Wigner-Seitz type of calculation (Wigner and Seitz 1933, 1934). This method has already had a considerable success in accounting for the properties of nuclear matter in a periodic model (Trainor, Bhatia, and Sood 1958). In the present context,  $\pi$ -mesons moving in Wigner-Seitz spheres of radius  $a$ , "see" a potential

$$(28) \quad V(r) = (-g/r) - [g/(2a-r)]$$

as shown in Fig. 2. By solving the appropriate Klein-Gordon radial equation using equation (28) with the usual condition of horizontal tangency at  $r = a$ ,

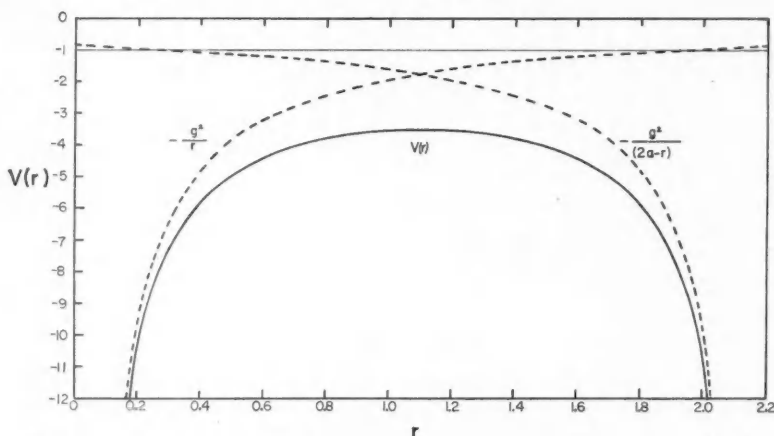


FIG. 2. Potential energy seen by the meson between two Wigner-Seitz spheres.  $V(r)$  is measured in units of the meson rest mass and the radial distance  $r$  is measured in units of  $10^{-13}$  cm. The volumes of the Wigner-Seitz spheres correspond to the volume per nucleon in nuclear matter.

one obtains an expression for the meson energy. By considering a range of values of the parameter  $a$ , one can arrive at an equilibrium radius which should agree with the observed value of the nuclear density.

These calculations will be the subject of a subsequent communication. However, we can anticipate one result at this point. Assuming that the detailed calculations yield a reasonable value of the Wigner-Seitz radius, viz.  $a = 1.1 \times 10^{-13}$  cm, equation (28) yields a value of the potential energy midway between nuclear sites of

$$(29) \quad -2g^2/a = -3.5mc^2.$$

Since this value exceeds the energy of the  $\pi$ -meson by a large factor, it seems probable that the mesons will be free to move from one core to another.

#### CONCLUSIONS

A direct attack by the methods of field theory appears to be a hopeless task for certain types of problems, e.g. the nature of the  $\pi$ -meson field in nuclear matter (Trainor and Sood 1959). This realization motivated the present work on the meson-atomic model of the nucleon. The work of Feld and Costa (Feld 1957, 1958; Costa and Feld 1960) suggests the importance of a model of the nucleon characterized by a  $\pi$ -meson moving in a  $p_{1/2}$  state of zero total energy. The question is how to achieve such a state on the basis of a simple model.

It has been shown in the present work that the Klein-Gordon equation for a single  $\pi$ -meson moving in an inverse square "hyper-Coulomb" field does possess formal solutions of the appropriate type. Specifically, if the hypercharge

$g$  is chosen sufficiently large so that  $g^2/\hbar c = \sqrt{2}$ , a state with a well-behaved wave function does exist, corresponding to angular momentum  $l = 1$  and energy eigenvalue  $\epsilon = 0$ . Moreover, no  $s$ -states occur which conserve energy so that (at least superficially) the  $p$ -state forms the ground state of the system. The model thus provides one with a "wave function" for the motion of the  $\pi$ -meson in the meson-atomic model.

Unfortunately, the strength of this field is so great as to render a single particle interpretation of the Klein-Gordon equation difficult from the point of view of field theory. Yet, even from this point of view, the result seems to have heuristic value. It raises the conjecture that, due to the repulsive nature of the centripetal effect,  $p$ -states retain an approximate single particle validity to much greater field strengths than do  $s$ -states. It is even possible that an exact treatment of the problem would reveal that the ground state could be reasonably approximated by a single particle  $p$ -wave.

An interesting interpretation of Sections III and IV is to regard the usual and unusual solutions as belonging to essentially different particles. For  $g^2/\hbar c = \sqrt{2}$ , one would then have the interpretation for the unusual solutions that only states with angular momentum  $l = 1$  were possible, either for the bound meson or for the meson in its free state. Photoproduction would then be restricted to the ejection of  $\pi$ -mesons in  $p$ -waves. For  $\pi$ -nucleon scattering one would also expect a strong  $p$ -wave resonance, while the screening of the core by the bound  $\pi$ -meson would allow the usual formulation of plane waves at infinity for the incident  $\pi$ -mesons.

Perhaps the main virtue of the meson-atomic model, in the form presented here, is in its purely suggestive value in attempting to understand something about the structure of nucleons in a simple way.

#### ACKNOWLEDGMENTS

The author is indebted to the University of Alberta Research Committee and to the National Research Council of Canada for financial assistance.

#### REFERENCES

- CASE, K. M. 1950. *Phys. Rev.* **80**, 797.
- CHEW, G. F. 1959. *Bull. Am. Phys. Soc.* **4**, 451.
- COSTA, G. and FELD, B. T. 1960. *Ann. Phys.* **9**, 354.
- FELD, B. T. 1957. *Ann. Phys.* **1**, 58.
- 1958. *Ann. Phys.* **4**, 189.
- FESHBACH, H. and VILLARS, F. 1958. *Revs. Modern Phys.* **30**, 24.
- LANDAU, L. D. and LIFSHITZ, E. M. 1958. *Quantum mechanics, non-relativistic theory* (Pergamon Press, London), p. 50.
- SCHIFF, L. I. 1955. *Quantum mechanics* (McGraw-Hill, New York), p. 321.
- TRAINOR, L. E. H. 1959. *Bull. Am. Phys. Soc.* **4**, 448.
- TRAINOR, L. E. H., BHATIA, A. B., and SOOD, P. C. 1958. *Nuclear Phys.* **9**, 287.
- TRAINOR, L. E. H. and SOOD, P. C. 1959. *Nuclear Phys.* **12**, No. 6, 625.
- WIGNER, E. and SEITZ, F. 1933. *Phys. Rev.* **43**, 804.
- 1934. *Phys. Rev.* **46**, 509.

## A LOW TEMPERATURE STANDARD OF TOTAL RADIATION<sup>1</sup>

R. E. BEDFORD

### ABSTRACT

A standard of total radiation used for the calibration of thermal radiation detectors is described. The black body source is a modified cone operating over the temperature range 40° C to 150° C; the detector is located in an isothermal (better than 0.01° C) environment. For the detectors used a decrease in sensitivity of approximately 1% during the first 24 hours was observed. This is probably due to the slow drying of the receiver surface by the gas used to displace the air. The day-to-day reproducibility of results is about 0.1%, and an over-all accuracy of 0.3% is claimed for radiation measurements made with the instrument. The scale of total radiation defined by four National Bureau of Standards carbon filament standards of thermal radiation probably differs from that defined here by -0.3%.

### INTRODUCTION

The measurement of radiant energy in absolute value is a problem of rather frequent occurrence in many fields of scientific research as, for example, radiometry, biology, chemistry, and agriculture. Such measurements are generally made by measuring the unknown radiation with a spectrally non-selective detector which has in turn been calibrated against that from a standard source. A black body\* is almost universally used as the standard source since both its spectral energy distribution and its total radiant intensity may be computed from a knowledge of its temperature and geometry. Incandescent lamps, which depend upon a previous comparison with a black body for their calibration, are sometimes used as secondary standards. The small carbon filament standards of thermal radiation issued by the National Bureau of Standards are examples of this latter type.

Modern requirements are such that the measurement of total radiation with an accuracy of better than 1% is often desired. Making total radiation measurements with an accuracy of a few tenths of one per cent places rather stringent demands on the calibration equipment. Coblenz (1933) expressed some doubts as to whether such an accuracy is in fact attainable. Leighton and Leighton (1932) obtained short-term precision of this order, but found that random fluctuations limited their over-all accuracy to about 1%. Guild (1937), on the other hand, feels he has established a scale of total radiation at the National Physical Laboratory with an accuracy of better than 0.5%. More recently Eppley and Karoli (1957) have obtained reproducibility of about 0.25% and an over-all accuracy of about 1%. Most total radiation measurements which have been reported were made in the period 1900 to 1925 for the purpose of determining experimentally the Stefan-Boltzmann constant, and

<sup>1</sup>Manuscript received June 27, 1960.

Contribution from the Division of Applied Physics, National Research Council, Ottawa, Canada.

Issued as N.R.C. No. 5901.

\*In this paper the term "black body" will be used in the sense of a good experimental approximation to a theoretical black body.

these do not agree to better than a few per cent (Coblentz 1921). In fact the best experimental value of  $\sigma$  differs by 1% from the theoretical value, which is usually assigned an uncertainty of about  $\pm 0.02\%$ .

The present paper describes an apparatus which has been constructed for the calibration of thermal detectors, and presents experimental data which indicate that an accuracy of a few parts in a thousand has been achieved.

#### THEORETICAL CONSIDERATIONS

In calibrating a detector, in our case a thermopile, the procedure consists basically of measuring its output when it is enclosed in a constant temperature surround and alternately exposed to a black body at some elevated temperature and to a shutter at the temperature of the surround. When the thermopile views the black body the net radiant flux density  $W_1$  at the thermopile is given by\*

$$(1) \quad W_1 = \sigma F_{01} \epsilon_1 \epsilon_0 (T_1^4 - T_{01}^4) \text{ watts cm}^{-2},$$

where  $\sigma$  is the Stefan-Boltzmann radiation constant,

$\epsilon_1$  is the emissivity of the black body,

$\epsilon_0$  is the emissivity of the receiving area of the thermopile,

$T_1$  is the temperature of the black body ( $^{\circ}\text{K}$ ),

$T_{01}$  is the temperature of the hot junctions of the thermopile ( $^{\circ}\text{K}$ ),

$F_{01}$  is a non-dimensional factor involving the geometry of the apparatus.

Equation (1) assumes that the thermopile is spectrally non-selective so that its emissivity for radiation characteristic of its own temperature is numerically equal to its absorptivity for radiation characteristic of the black body temperature. Similarly, when the thermopile views the shutter, we get

$$(2) \quad W_2 = \sigma F_{02} \epsilon_2 \epsilon_0 (T_2^4 - T_{02}^4).$$

In our case the apparatus is designed so that  $F_{01} \equiv F_{02}$ . Hence subtraction of equation (2) from equation (1) yields

$$(3) \quad W_1 - W_2 = \sigma F_{01} \epsilon_0 [\epsilon_1 (T_1^4 - T_{01}^4) - \epsilon_2 (T_2^4 - T_{02}^4)].$$

We might note that if the surroundings change temperature when the shutter is opened and closed another term of the form  $\sigma F_{03} \epsilon_0 \epsilon_3 (T_{31}^4 - T_{32}^4)$  should be added to the right-hand side of equation (3). Ideally  $T_2 = T_{02}$ ,  $T_{31} = T_{32}$ , and equation (3) reduces to

$$(4) \quad W_1 - W_2 = \sigma F_{01} \epsilon_0 \epsilon_1 (T_1^4 - T_{01}^4).$$

If  $V_1$ ,  $V_2$  are the e.m.f.'s developed by the thermopile when exposed to the black body and shutter respectively, then the sensitivity (or responsivity)  $S'$  of the thermopile is defined as

$$S' = \epsilon_0 [(V_1 - V_2) / (W_1 - W_2)],$$

\*In the following equations the subscripts 0, 1, 2, 3 refer to the thermopile, black body, shutter, and surround respectively. Double subscripts on the temperatures are interpreted thus:  $T_{02}$  is the temperature of the receiving area of the thermopile when exposed to radiation from the shutter, etc.

the quantity  $\epsilon_0$  appearing since the sensitivity is to be defined in terms of the net incident radiant flux density, and not the absorbed radiant flux. Substituting from equation (4):

$$(5) \quad S' = (V_1 - V_2) / \sigma F_{01} \epsilon_1 (T_1^4 - T_{01}^4).$$

The temperature of the thermopile hot junctions ( $T_{01}$ ) appearing in this expression is dependent on  $T_1$  and so makes computations tedious. If the thermopile response is linear, we may replace  $T_{01}$  by the temperature of the cold junctions ( $T_0$ )\* and we then have

$$(6) \quad S = (V_1 - V_2) / \sigma F_{01} \epsilon_1 (T_1^4 - T_0^4),$$

where  $S$  differs from  $S'$  by only 0.1%.  $S$  is the quantity commonly used in radiation calculations.

Equations (1)–(6) implicitly assume that there is no absorption of radiation by the medium between the source and the thermopile.

We must now consider the quantities  $F_{01}$  and  $\epsilon_1$  in equation (6). The quantity  $F_{01}$  is dependent on the geometry of the apparatus and may be written  $F_{01} = f_{10}(A_1/A_0)$ , where  $A_1$  and  $A_0$  are the effective emitting and receiving areas respectively, and  $f_{10}$  (variously called the geometrical factor, view factor, angle factor, shape factor, or configuration factor) is the fraction of the total energy emitted by the source which is incident upon the thermopile. Expressions for  $f_{10}$  for many configurations have been compiled by various authors. In the limiting case of a small source,  $F_{01} = \Omega/\pi$ , where  $\Omega$  is the solid angle subtended by the source at the thermopile. In the apparatus described here a circular aperture of radius  $a_1$  in front of the black body opening defines the solid angle seen by the thermopile, and acts as a virtual source at the temperature  $T_1$ . No aperture was placed before the thermopile and hence the circular hot junction of radius  $a_0$  itself defines the receiving area. For this configuration of two plane, parallel, circular surfaces with a common central normal and separated by a distance  $l$ , the expression for  $F_{01}$  is (see, for example, Jakob 1957)

$$(7) \quad F_{01} = [(a_1^2 + a_0^2 + l^2) - \sqrt{(a_1^2 + a_0^2 + l^2)^2 - 4a_1^2 a_0^2}] / 2a_0^2.$$

Since in our case  $a_1 a_0 \sim 10^{-3} l^2$ , we may neglect second and higher order terms in the expansion of the square root and write

$$(8) \quad F_{01} = a_1^2 / (a_0^2 + a_1^2 + l^2).$$

If the axes of the thermopile and the source are parallel but not colinear, being separated by a distance  $x$ , and if  $a_0$  is negligibly small compared to  $l$ , then the geometrical factor becomes (Jakob 1957)†

\*In equation (5) the term  $(T_1^4 - T_{01}^4)$  may be rewritten as  $[(T_1^4 - T_0^4) - (T_{01}^4 - T_0^4)]$ .

For a linear thermopile response the second term in parentheses is proportional to the first (this has been experimentally verified). Equation (5) may then be written

$$S' = (V_1 - V_2) / k \sigma F_{01} \epsilon_1 (T_1^4 - T_0^4)$$

where  $k$  is a constant. In our case  $k$  has the approximate value 0.999.

†A typographical error occurs in the expression for  $F_{01}'$  as given in the reference.



$$(9) \quad F_{01}' = \frac{1}{2} \left[ 1 - \frac{l^2 + x^2 - a_1^2}{\sqrt{x^4 + 2x^2(l^2 - a_1^2) + (l^2 + a_1^2)^2}} \right].$$

While equation (7) holds quite generally, equation (9) is valid only when the dimensions of the receiving area are very much smaller than the distance from the thermopile to the source.

It remains to evaluate  $\epsilon_1$ , the emissivity of the black body. Three methods are available, none of which is wholly satisfactory because of lack of experimental confirmation. Assuming that the walls of the cavity are perfectly diffusely reflecting, Buckley (1934) developed an expression for the emissivity in the form of an integral equation. However, even for simple geometrical shapes, the integration is laborious and for a cavity such as is used here this method is not feasible. Gouffé (1945) introduced the further assumption that after two reflections the distribution of reflected radiation in the cavity is uniform, and developed a closed expression for the emissivity. His equation is

$$(10) \quad \epsilon_1 = \frac{\epsilon \{1 + (1 - \epsilon)[(n/N) - (n/N_0)]\}}{\epsilon[1 - (n/N)] + (n/N)},$$

in which  $\epsilon$  is the emissivity of the material of the walls of the black body,  $n$  is the area of the opening ( $n \ll N$ ),  $N$  is the total surface area of the black body (including the opening), and  $N_0$  is the area of the surface of a sphere of diameter equal to the depth of the black body in the direction normal to the opening  $n$ .

A third and more exact means of obtaining a value for  $\epsilon_1$  is due to De Vos (1954), who takes some account of the nature of the reflection from the cavity walls. Edwards (1956) has applied this method to the case of a conical black body for several different types of reflection.

The only check of these emissivity calculations appears to be that of Vollmer (1957). He attempted to test experimentally the validity of Buckley's method for shallow cylinders and concluded that theory and experiment were in agreement within his limits of error (a few per cent).

#### DETAILS OF THE APPARATUS

The main components of the apparatus are the black body source, the radiation detector, and the radiometer housing, which provides a constant temperature surround for the radiation detector, defines the aperture of the system, and contains the shutter.

##### 1. Black Body Source

###### Construction

The black body cavity consists of a right circular cone with a cylindrical extension at its base, immersed in a stirred liquid bath, and is shown on the right-hand side in Fig. 1. The cone itself is 56 cm long with an aperture of 17.8 cm, giving an angle for the cone of about  $18^\circ$ . The cylindrical portion attached to the base of the cone has a length of 18.2 cm and is partially closed at the end by a copper plate having a bevelled aperture of diameter 12.7 cm. The portion of the cavity actually seen by the thermopile is restricted to the

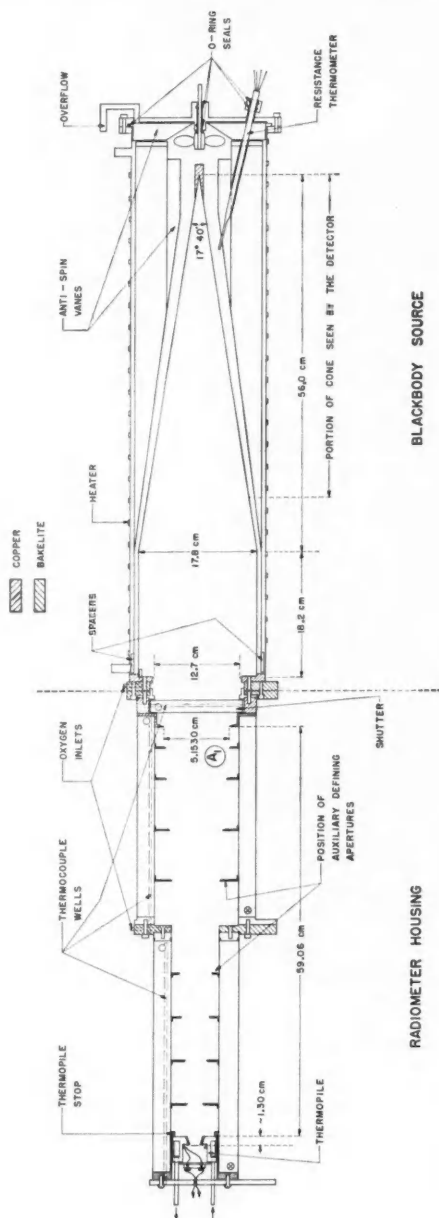


FIG. 1. Black body source and radiometer housing.

inner 50 cm of the cone. The cavity is constructed from 0.10-cm copper sheet, the cone being rolled to give a butt joint at the seam. To ensure a sharp point at the tip, the end of the cone consists of a conical hole drilled in a copper block which has then been fitted flush with the rolled portion. The radiator is blackened inside with commercial flat black paint.

The brass tank in which the radiator sits is filled with oil or water which is circulated by a propeller. The liquid flows from the apex to the base of the cavity and then returns along the outer wall of the tank; appropriate antispin vanes direct the flow and prevent vortex formation at the propeller. The 10-ohm heater element of chromel-A ribbon is wound along the length of the outer wall, and the heat input is supplied from a variac capable of delivering a maximum power of 2 kilowatts. The whole unit is thermally insulated with an outer wrapping of Fiberglas. Servicing or refilling may be readily accomplished by removing the end plate to which the propeller is attached. An overflow tap is provided which allows the bath to be kept completely filled at all temperatures.

#### *Temperature Measurement*

The temperature region over which the black body operates is roughly 40° C to 150° C, and since rather accurate measurements of these relatively low temperatures are required, a platinum capsule resistance thermometer, inserted in the liquid bath near the conical surface, is used. Its resistance is measured by comparing it with a standard variable resistor\* using an isolating potential comparator (Dauphinee 1953), a method which is particularly convenient in this instance because it allows the black body temperature to be displayed on a recorder, and so any fluctuations during the course of a radiation measurement are easily detected.  $T_1$  may be kept constant to within 0.01° C at the lower temperatures and 0.03° C at the higher temperatures by using simple manual control with the variac.

Since the resistance thermometer measures the temperature of the liquid in the bath, a differential copper-constantan thermocouple, soldered in various locations, was used to determine any temperature differences between the liquid and the emitting surface of the black body, and also to determine any temperature differences along this surface. Measurements were made with the cone in both the horizontal and vertical (base down) positions and with the temperature at about 120° C. Except for very slow propeller speeds, gradients in the liquid itself did not exceed 0.01° C. Although in the horizontal position the thermocouple indicated that the surface was cooler than the liquid by as much as 1° C at the open end, in the vertical position measured temperature differences did not exceed 0.01° C in the cone, and were not greater than 0.1° C at the extreme open end. The thermocouple, of course, does not measure any temperature drop across the paint, but calculation shows this to be negligible. The accuracy of the measurements is dependent on the heat losses from the junction through the thermocouple wires, which cause the observed temperature

\*A Cambridge vernier potentiometer is used as the standard four-terminal variable resistor. This gives least resistance steps of 0.00045 ohm, corresponding to steps of approximately 0.0045° C.

to be, if anything, lower than the true surface temperature. In any case, since the detector sees only the cone, and since the source is always used vertically, the observed temperature differences are not significant.

#### *Emissivity*

The reflectivity in the visible spectrum of the black paint used on the walls of the black body was measured and found to be about 0.05, and since most black paints have a somewhat higher reflectivity in the infrared than in the visible, the emissivity ( $\epsilon$ ) of the walls is assumed to be 0.85–0.90 for the wavelength region involved. Using this value and the dimensions given above, we obtain from equation (10) a value of 0.998 for the emissivity ( $\epsilon_1$ ) of the black body. The value of  $\epsilon_1$  is rather insensitive to changes in  $\epsilon$ , varying from 0.996 to 0.999 as  $\epsilon$  changes from 0.80 to 0.95. This value of  $\epsilon_1$  assumes diffuse reflection in the cavity. Most sources (see, for example, De Vos 1954) indicate that the result of Gouffé's calculation (equation (10)) is, if anything, too low. A second estimate of the emissivity may be made from the calculations given by Edwards (1956) for a cone. His figures indicate that the emissivity increases as the reflection becomes more specular. For a cone with a ratio of length to radius of aperture of 10, and for partially specular reflection with  $\epsilon \sim 0.4$ , Edwards finds  $\epsilon_1 = 0.998$ . Our black body is probably at least as black as a cone with the above ratio of 10, and since in our case  $\epsilon \sim 0.85$ , then  $\epsilon_1$  should be at least 0.998. Hence if we take  $\epsilon_1 = 0.998$  we should not be in error by more than  $\pm 0.1\%$ .

#### *2. Radiometer Housing*

The radiometer housing consists essentially of two concentric cylindrical tubes as shown on the left-hand side in Fig. 1. The inner tube is of copper, the outer of brass, and temperature-controlled water is circulated between them. The unit has an external wrapping of Fiberglas. For convenience in use the housing is made in two sections which may be bolted together. The second section (containing the detector) may then be removed and used as a separate unit when so desired.

The circular aperture  $A_1$ , which defines the solid angle subtended at the thermopile by the black body, is located immediately in front of the shutter, and is a part of the same copper end plate which closes the water channel. Thus there is excellent thermal contact with the cooling water. The diameter of the aperture is nominally 10 cm, it is distant 60 cm from the receiver, and its bevelled edge is machined to a thickness not greater than 0.0025 cm. Along the length of the housing there are eight radiation baffles. These take the form of circular disks with central apertures of progressively increasing radii and are so spaced that the thermopile can see none of the cylinder wall, and essentially only the first of the baffles. The solid angle defined by these baffles is only slightly larger than that determined by the limiting aperture  $A_1$ , so that the thermopile is able to see an annulus on the face of  $A_1$  less than 0.1 cm in width. The baffles are constructed from 0.16-cm copper, their bevelled edges have a thickness not greater than 0.0025 cm, and they are soldered in place to ensure good thermal contact.

The shutter, the faces of which are made from 0.16-cm copper, slides in

ways situated between the aperture  $A_1$  and the black body. When the shutter is open, the thermopile sees through a tapered hole in it into the black body; when it is closed the thermopile views that part of its plane face defined by the aperture  $A_1$ . The whole inside of the housing and both faces of the shutter are blackened with commercial flat black paint. The blackening of the side of the shutter next to the source results in roughly equal heat losses from the source when the shutter is opened and closed; this is an aid in maintaining the source temperature constant.

When mounted in position the radiometer housing is fastened directly to the black body by the bakelite supports as shown in Fig. 1. No difficulty was experienced in properly aligning the system.

In general the thermopile and radiometer housing are maintained near room temperature. The cooling water is circulated from a small reservoir through the shutter, the first section of the housing, the thermopile case (described below), and the second section of the housing in series in that order by means of a standard gear pump, the flow being about 150 cc per second. A portion of the return line is immersed in a refrigerated bath which may be maintained at any desired temperature within an accuracy of  $0.1^\circ\text{C}$ , and a heater coil is installed in the return line where it enters the reservoir. A mercury thermoregulator having an adjustable set point and a sensitivity of  $\pm 0.005^\circ\text{C}$  is placed in the reservoir, and controls the operation of the heater. The room in which the apparatus is housed is both temperature- and humidity-controlled, and it is possible to control the temperatures of the various components to better than  $0.01^\circ\text{C}$ . In order to measure temperatures over the housing, thermocouple wells of German silver tubing are installed in the shutter, one near the water inlet and one at the outlet, and in each of the cylindrical sections. The measurement of temperatures and gradients will be described in the next section.

Absorption of radiation by carbon dioxide and water vapor is eliminated by displacing the air in the housing and black body by a dry, non-absorbing gas. Oxygen is particularly convenient to use in our case.

### 3. Detector

The only radiation detectors used in the present study were 8- and 16-junction bismuth-silver thermopiles.\* The receiving areas are circular, approximately 1 cm in diameter, and formed from silver foil blackened with either gold black or carbon black.

Each thermopile is permanently mounted in a water-cooled brass case of the form shown in Fig. 2. The receiving area is centered behind the tapered opening, the "staircase type" surface of which is designed to prevent reflection of radiation from the case to the receiving area. The front of the case, the walls of the opening, and the whole of the interior are coated with flat black paint. A German silver thermocouple well is set into the water jacket. No window is used on the thermopile. This unit may be mounted in the back of the radiometer housing, its position being determined quite reproducibly by a stop as shown in Fig. 1.

\*The thermopiles were obtained from the Eppley Laboratory, Inc., Newport, R.I.

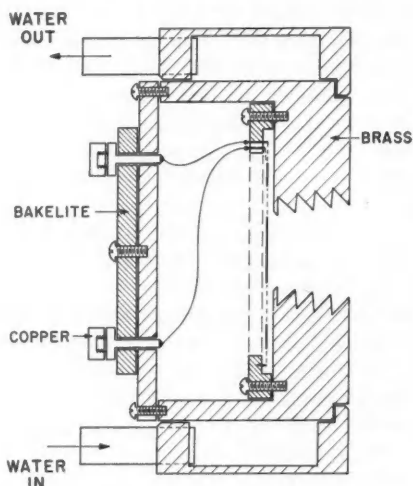


FIG. 2. Thermopile case.

#### 4. Instrumentation

For black body temperatures in the range  $40^{\circ}$  to  $150^{\circ}$  C the e.m.f. developed by the thermopile varies from about 10 to 200 microvolts, and for an accuracy of 0.1% it is necessary to measure differences in e.m.f. with accuracies of up to 0.01 microvolt ( $\mu$ v). To accomplish this the output of the thermopile is amplified by a photocell galvanometer amplifier, and the resulting signal measured with a Cambridge vernier potentiometer giving least steps of 0.1  $\mu$ v. In place of a galvanometer a Leeds and Northrup d-c. microvolt amplifier and recording potentiometer are used with the vernier potentiometer. With this arrangement the thermopile output can be continuously displayed, and an average reading obtained over any desired time interval. Small deviations of the e.m.f. from balance are read from the recorder chart.

The accuracy of measurement of the thermopile output is dependent on the characteristics of the galvanometer amplifier, the circuit of which is shown in Fig. 3. The resistor  $R_0$  and the feedback resistor  $R_f$  are oil-immersed precision decade-resistance units.  $R_0$  may be varied from 100 to 1000 ohms in 100-ohm steps, and  $R_f$  from 1 to 100 ohms in 1-ohm steps. If the thermopile produces an e.m.f.  $V$  and the voltage developed across  $R_0$  is  $V_0$ , it may readily be shown that

$$(11) \quad V = V_0 \frac{R_f}{R_0} \left( 1 + \frac{R_G + R_T + R_f}{MR_f} \right)$$

where  $R_T$  is the resistance of the thermopile,  $R_G$  is the resistance of the galvanometer, and  $M$  is the gain of the system without feedback. In our case  $M \sim 2300$ ,  $R_G = 10$  ohm, and  $R_T \sim 10$  ohm (for an eight-junction thermopile), so that  $V$  is given approximately by

$$(12) \quad V \sim V_0 (R_f/R_0).$$

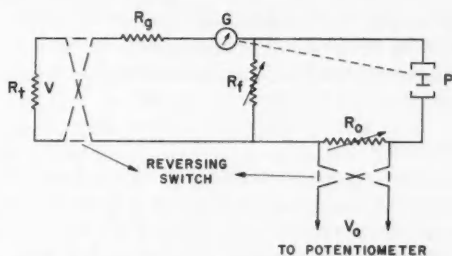


FIG. 3. Circuit diagram of the galvanometer amplifier. Light from the mirror of the galvanometer, G, falls on two barrier layer cells, P, connected in opposition.

The amplification thus varies approximately linearly with the feedback resistance, and small departures from linearity may be calculated using equation (11). An amplification of 10 was used almost exclusively ( $R_f = 50$  ohm,  $R_0 = 500$  ohm).

Tests have shown the amplification to be linear with respect to the input signal to within 0.01%, and to be independent of lamp voltage fluctuations, for all conditions of the experiment.

The limit of measurement set by the noise generated in the amplifier itself is of the order of  $0.003 \mu\text{V}$ . This is much smaller than the limit set by the fluctuations in thermopile output (due to adiabatic pressure changes, etc.), of about  $\pm 0.01 \mu\text{V}$  on calm days rising to  $\pm 0.05 \mu\text{V}$  on windy days. In general, the thermopile e.m.f. can be measured with an uncertainty of  $\pm 0.02 \mu\text{V}$ .

## EXPERIMENTAL

### 1. Geometry

In order to calculate the quantity  $F_{01}$  we must determine the quantities  $l$ ,  $a_1$ ,  $a_0$  of equation (7). The distance between the plane of the aperture  $A_1$  and the receiving surface of the thermopile ( $l$ ) is the sum of two parts—the distance from  $A_1$  to the face of the thermopile case ( $b_1$ ) and the distance from the front of the case to the receiving surface ( $b_2$ ). The length  $b_1$  was measured with a micrometer depth gauge for repeated opening and reassembling of the radiometer housing, and a value  $b_1 = 59.06 \pm 0.015$  cm was obtained. The distance  $b_2$  varies from thermopile to thermopile. It was measured for each thermopile with a graduated microscope by focussing alternately on the front of the case and on the receiver, and values of  $b_2$  of the order of 1.30 cm were obtained with an error of less than  $\pm 0.01$  cm. Measurements of  $b_2$  were made with the thermopiles in both the horizontal and vertical positions, and no systematic dependence of  $b_2$  on position was obtained. Hence the aperture to receiver distance ( $b_1 + b_2$ ) is of the order of 60.36 cm with an error not greater than  $\pm 0.025$  cm.

The mean diameter  $2a_1$  of the defining aperture  $A_1$  was obtained from measurements of four diameters at approximately  $45^\circ$  intervals using a precision micrometer slide and was found to be  $10.306 \pm 0.001$  cm. The diameters of the receiving areas of the thermopiles were similarly determined, and these varied from 0.966 to  $0.976 \pm 0.001$  cm.

From these dimensions  $F_{01}$  is calculated to be of the order of 0.007235 and hence the radiant flux density at the thermopile varies from roughly 100 to 1000  $\mu\text{W}$  per  $\text{cm}^2$  for black body temperatures in the range  $40^\circ$  to  $150^\circ\text{C}$ .

## 2. Testing

### *Temperature Gradients*

In calibrating the thermopile we subtract its reading when the shutter is closed from its reading when the shutter is open and make use of equation (6). For the equation to apply we must be certain that the only additional radiation incident on the thermopile when the shutter is open does in fact come from the source, and that no significant amount arises from reflections or from changing temperature distributions in the housing. With the system mounted horizontally considerable trouble was caused by convection losses from the black body. Temperatures changed appreciably in the housing when the shutter was opened, the thermopile output showed considerable drift, and the black body temperature dropped. To eliminate these effects the apparatus is mounted on a pivoted board so that it can be turned vertically, with the source uppermost, when measurements are made.

Temperature changes in the radiometer housing were measured with differential copper-constantan thermocouples inserted in the various wells mentioned earlier. With the black body at room temperature no differences greater than a few thousandths of a degree were observed. As the black body is heated to  $150^\circ\text{C}$  the temperature of the housing as a whole rises about  $0.05^\circ\text{C}$ . At these higher temperatures with the shutter closed, the shutter is cooler than the first cylindrical section by about  $0.01^\circ\text{C}$ , which in turn is cooler than the second section by about  $0.01^\circ\text{C}$ . There is no observable temperature difference between the second section (containing the thermopile) and the thermopile itself. When the shutter is opened the temperature of the first section rises at most  $0.003^\circ\text{C}$  while the second section and thermopile, maintaining their equality of temperature, change less than  $0.002^\circ\text{C}$ .

In order to show that the temperatures of the inner wall and of the baffles are not significantly different from that of the water, chromel-alumel thermocouples were soldered to  $A_1$ , to each baffle of the first section, and to the cylinder wall, all being shielded from direct radiation from the source. Changes in temperature were recorded as the shutter was opened for source temperatures up to  $135^\circ\text{C}$ . At the highest temperature, the thermocouple on the wall indicated a difference of  $0.01^\circ\text{C}$ , about three times that observed in the water itself. The thermocouples on the baffles furthest from the source changed in temperature about  $0.02^\circ\text{C}$ , while those on the closer ones, and on  $A_1$  itself, changed by at most  $0.10^\circ\text{C}$ . Again we must remark that these figures are subject to the accuracy with which the thermocouples register the true surface temperature. The observed changes will be, if anything, too great.

Thermocouples in the second section near the thermopile, even when in poor thermal contact with the surface, were not affected by the opening and closing of the shutter.



A continuous record of the temperatures in the housing is kept by means of four copper-constantan thermocouples in the German silver wells. The output of one thermocouple is measured with a Cambridge vernier potentiometer, and those of the others are compared with it by means of an apparatus for the automatic comparison of thermocouples (Dauphinee 1955). All four can thus be simultaneously displayed on a recorder. The thermocouples were calibrated to better than  $0.01^{\circ}\text{C}$  against a platinum resistance thermometer. The temperature as recorded by the thermocouple in the thermopile case or in the second section of the housing gives  $T_0$  for use in equation (6).

The thermocouples in the various wells show the cycle of the thermostat, the peak-to-peak change in amplitude being about  $0.01^{\circ}\text{C}$  in the shutter, and  $0.002^{\circ}\text{C}$  in the housing and thermopile case. This appears on the thermopile output as a cycle with a peak-to-peak change in amplitude of from 0.10 to  $0.25\text{ }\mu\text{v}$ . This relatively large variation in the e.m.f. is due, of course, to the different rates of response of the hot and cold thermopile junctions to temperature changes in the case. The variation is quite regular and the mean e.m.f. is easily read to within  $0.02\text{ }\mu\text{v}$ .

#### *Reflections*

The problem of estimating the contribution of reflected radiation is rather difficult. To make an evaluation, an aperture,  $A_2$ , was cut from 0.64-cm brass such that it could be put into the first section of the housing and held flush against the back baffle (the position is indicated in Fig. 1). The size of this aperture was such that when in position it formed the limiting aperture for the system. The radiant flux density at the thermopile in this case could easily be calculated from the dimensions, just as for  $A_1$ . The temperature rise of  $A_2$  itself was measured by a thermocouple set into a hole drilled in it, and this rise (of the order of  $0.10^{\circ}$  to  $0.20^{\circ}\text{C}$ ) was accounted for in calculating the flux density. The sensitivity of a thermopile was then measured with apertures  $A_1$  and  $A_2$  for several black body temperatures both with and without black paint in the housing. For  $A_2$ , removal of all the paint produced no change in the observed thermopile sensitivity, indicating that reflections in the portion between the detector and  $A_2$  were not significant. However, an increase of 5% occurred in the sensitivity observed with  $A_1$  when the paint was removed. Since at this time the solid angle defined by the baffles was large enough that the thermopile could see almost the whole face of  $A_1$ , it seemed likely that most of this reflected radiation was coming from the unpainted face of  $A_1$ . The baffles were repositioned so that the receiver could see only a small annulus on  $A_1$ , and measurements repeated. The apparent change in sensitivity caused by removing the paint was reduced to about 2%. With the housing painted the sensitivity with  $A_1$  was 0.3% higher than with  $A_2$ . Three different black paints with rather different angular reflection characteristics but probably about the same total reflection coefficient produced the same result. The baffles were then machined so that their edge thicknesses were reduced from their original values of 0.02 to 0.05 cm to not more than 0.0025 cm. Within the accuracy of measurement the discrepancy between sensitivities

with  $A_1$  and  $A_2$  disappeared. Measurements were also made with the face of  $A_2$  painted and unpainted, but no change of reading resulted.

Reflected radiation arising from the thermopile case itself was also evaluated. Measurements were first made with a case without the "staircase wall", and with a series of apertures over the opening which progressively exposed more of the wall to the incident radiation. These measurements indicated that reflections from the wall contributed more than 20% of the observed signal, and that the amount of reflected radiation was proportional to the length of the wall visible from the radiator. Similar measurements were then made with a case with the "staircase wall". Because of the geometry it was impossible in this case to shield the whole wall without at the same time limiting the aperture of the system. However, the reading with all of the wall exposed was only about 0.5% higher than the reading with roughly half of the wall covered. Hence the reflected radiation is probably about 1% of the total. Since the thermopile is always used in the same case, this small amount of reflected radiation will be relatively constant independent of the nature of the source and so can be absorbed in the constant of calibration.

#### *Defining Apertures*

The effect of using different defining apertures was also examined. Four circular apertures similar to  $A_2$ , but with different size openings were used at the position of  $A_2$ . Another was made which could be placed against the first baffle at the end of the second section of housing (see Fig. 1). Diameters and distances were measured as for  $A_1$ , but the relative errors were larger because of the smaller dimensions. For all the apertures, including  $A_1$ , the displacement ( $x$ ) of the center of the thermopile relative to the axis of the aperture was measured and  $F_{01}'$  calculated according to equation (9). While for  $A_1$  the difference between  $F_{01}$  and  $F_{01}'$  was not significant (0.005%), the displacement was as high as 0.7 cm for some of the other apertures, corresponding to a difference between  $F_{01}'$  and  $F_{01}$  of up to 0.2%. The temperature rises of the auxiliary apertures were measured and accounted for. With these corrections applied, the thermopile sensitivities obtained with all the apertures agreed within the limits of experimental error. Thus there was no systematic dependence of the calibration on either the distance between source and receiver or the magnitude of the solid angle subtended by the source.

#### *Absorption*

For all the measurements the air in the housing and in the black body is displaced by dry oxygen (or some other dry gas). The oxygen is flushed through for several minutes at a rather high pressure to drive out any trapped air, and then a very slight overpressure is sufficient to prevent re-entry. The efficiency of this exchange process was examined by means of an Aminco-Dunmore electric hygrometer multiple-range sensing element. The hygrometer was placed at various locations in the system and the rate of decrease of relative humidity noted as the oxygen entered. Even with the meter placed in a polyethylene bag strapped to the end of the housing (thermopile removed)

and with a low oxygen flow, the meter indicated a relative humidity of zero in less than 2 minutes. Only a trickle of oxygen was required to maintain these conditions. The effect of the oxygen flow on the thermopile reading was also examined. High flow rates cause fluctuations and drift of the output, but low constant flow rates do not affect the output. Sudden changes in oxygen flow are seen by the thermopile because of the adiabatic change of pressure.

### 3. Calibration

When a thermopile is calibrated with the black body standard, it is mounted in the radiometer housing and allowed to stand for at least 48 hours in the atmosphere of dry oxygen (for reasons discussed later). With the cooling water circulating and the shutter closed, the equilibrium "zero" output and the cold junction temperature of the receiver are displayed while the source is being heated to the desired temperature. Once the black body has reached an equilibrium, the shutter is opened and the new thermopile output recorded. When a good average reading has been obtained (in 2 to 5 minutes), the shutter is closed again and the zero once more shown. The cycle is repeated as often as desired for any one black body temperature, usually three to five readings being sufficient. The procedure is then repeated for different source temperatures. Values of the thermopile sensitivity are computed from equation (6). In this manner measurements of the sensitivities of the thermopiles were carried out over a period of several months for source temperatures between 40° and 150° C.

The change of sensitivity of the detector with the temperature of the cold junctions was determined by altering the set point of the thermostat and recalibrating. These measurements were made over a range of about 10° C near room temperature using a black body temperature of about 115° C.

Since in practice the thermopiles are not always used in an atmosphere of oxygen, and since the sensitivity is a function of the thermal conductivity of the surrounding medium, some measurements were made with different gases in the system, notably nitrogen, helium, and argon. One thermopile was also mounted in a vacuum enclosure and its reading as a function of the pressure determined for a number of gases using as a source an incandescent lamp operated from a regulated power supply.

The scale of total radiation defined by the present apparatus was compared with that defined by four carbon filament standards of thermal radiation purchased from the National Bureau of Standards. The lamps were set up according to the instructions issued with them and the thermopiles calibrated in an atmosphere of oxygen with each lamp for three different lamp currents. The lamps were operated from a regulated d-c. power supply, and currents and voltages were measured using standard resistors and potentiometers. For these calibrations only the second section of the housing was used to enclose the thermopile, and the recommended shield and shutter made from air-separated sheets of cardboard were used. The detector housing was temperature-controlled in the usual way.

## ERRORS

In order to assess the over-all accuracy of the radiation measurements we list here the various errors involved, the percentage figures referring to uncertainties in the computed value of  $S$ .

1. An error of  $0.01^\circ \text{C}$  in  $T_1$  will cause errors of 0.06% and 0.01% in the calculated radiant flux densities at  $40^\circ \text{C}$  and  $150^\circ \text{C}$  respectively. The thermometer itself contributes no significant error, but the fluctuations in  $T_1$  noted earlier contribute an uncertainty of about  $\pm 0.04\%$ .

2. An error of  $0.01^\circ \text{C}$  in  $T_0$  will cause errors of 0.05% and 0.004% in the calculated radiant flux densities for  $T_1$  at  $40^\circ \text{C}$  and  $150^\circ \text{C}$  respectively. Fluctuations in  $T_0$  were less than  $0.01^\circ$  but the uncertainty of thermocouple calibration was  $\pm 0.01^\circ \text{C}$ . Hence an uncertainty of the order of  $\pm 0.01\%$  results from this limitation.

3. As noted earlier the accuracy of measurement of the thermopile output is variable. The uncertainty may be as high as  $\pm 0.05 \mu\text{V}$  in  $10 \mu\text{V}$  (that is  $\pm 0.5\%$ ) or as low as  $\pm 0.01 \mu\text{V}$  in  $100 \mu\text{V}$  (that is  $\pm 0.01\%$ ). On the average it is better than  $\pm 0.05\%$ .

4. The accuracy of the calculated value of  $\epsilon_1$  was discussed earlier and an uncertainty of  $\pm 0.1\%$  was given.

5. The estimated uncertainties in the quantities  $l$ ,  $a_1$ ,  $a_0$  lead to an uncertainty of  $\pm 0.09\%$  in the calculated value of  $F_{01}$ .

6. Errors due to temperature changes in the radiometer housing are as follows.

Since the shutter and thermopile temperatures differ by at most  $0.02^\circ \text{C}$  for the larger values of  $T_1$ , that is  $(T_{02} - T_2) \leq 0.02^\circ \text{C}$ , the second term in parentheses in equation (3) contributes at most 0.014% to the radiant flux density. From the geometry of the apparatus and using an equation similar to (1) we may compute the approximate radiant flux density at the thermopile due to temperature changes in the various components. Thus for a temperature change  $\Delta T$  relative to the thermopile we get respectively 2.2  $\Delta T$  and 130  $\Delta T \mu\text{W}$  per  $\text{cm}^2$  contributed from the first and second sections of the housing. Using the observed temperature changes discussed earlier, we get as a maximum possible contribution from the first section ( $\Delta T = 0.10^\circ \text{C}$  at  $T_1 = 100^\circ \text{C}$ ) 0.04%. A more probable figure is 0.013% (using  $\Delta T = 0.03^\circ \text{C}$ ). No temperature change was observed in the second section, but changes of less than  $0.001^\circ \text{C}$  would not have been detected. Hence the second section contributes less than 0.02% (assuming a maximum  $\Delta T$  of  $0.001^\circ \text{C}$  at  $T_1 = 100^\circ \text{C}$ ). These values are all negative since they can only increase the radiant flux density and so decrease the observed value of the thermopile sensitivity.

7. The contribution to the radiant flux density of reflected radiation in the radiometer housing is at most 0.1%. (Although no systematic error was detected, the accuracy of measurement places this limit.)

8. Negligible errors arise from gradients in the black body, absorption of radiation along the path between the black body and the detector, non-linear effects of the galvanometer amplifier, and instrumentation.

9. If the blackening material of the thermopile receiving surface were spectrally selective, systematic differences in the sensitivity observed for different black body temperatures would appear. With the exception of the results from one damaged thermopile (discussed later), no such effect was observed.

Since the errors considered in 6 and 7 are additive and total to a maximum of  $-0.15\%$ , all values of the thermopile sensitivity,  $S$ , are reduced by  $0.1\%$  from the values directly calculated from equation (6). The uncertainty of this correction is at most  $\pm 0.05\%$ .

Of the errors listed, those in 1-3 are essentially random and so should tend to average over several determinations (usually three to five are made at any one black body temperature as already pointed out). The final uncertainties given have, to some extent, taken this into account. The factors involved in 4-9 represent systematic uncertainties in the results.

The sum of the maximum uncertainties is  $0.34\%$ . It seems reasonable to assume therefore that values of total radiation obtained with the standard from the average of several determinations are accurate to within  $\pm 0.3\%$ .

#### RESULTS AND DISCUSSION

The results of the calibration of five thermopiles against the black body radiation standard are shown in Fig. 4, where the sensitivity  $S$  as obtained from equation (6) and corrected as noted above is plotted as a function of the temperature of the source,  $T_1$ . The value of the Stefan-Boltzmann constant  $\sigma$  used to compute  $S$  was that given by Cohen *et al.* (1955),  $\sigma = 5.669 \times 10^{-12}$  watt  $\text{cm}^{-2}$   $\text{deg}^{-4}$ . Each point plotted represents the mean of from two to five successive individual determinations as described above, and the vertical lines through the points indicate the total spread of these determinations. The scatter is dependent on atmospheric conditions; on windy days more noise is generated by the thermopile and the precision is reduced. However no time-dependent systematic changes of the results were observed. In the figure the total range of the points for each thermopile is about  $0.4\%$ . The mean sensitivity of each thermopile is given, together with the standard deviation of an individual point. The standard deviations have values of about  $\pm 0.12\%$ , and were computed with respect to the mean, and not with respect to the least square line.

The straight lines drawn through the points are least squares fits to the data and ideally they should be horizontal. Only thermopile No. 4 shows any significant systematic dependence of  $S$  on  $T_1$  and for it a rise in  $S$  of  $0.5\%$  occurs between  $30^\circ$  and  $150^\circ$  C. (A better fit to this data could be obtained by a second degree curve with a minimum at about  $70^\circ$  C.) Such a sensitivity change with  $T_1$  could be caused by non-uniform absorption by the receiver or the black body, systematic errors in the measurement of  $T_1$  or  $T_0$ , non-linearity of the thermopile response, or extraneous radiation arising from reflections and temperature changes in the housing. The effect cannot be attributed to the source or housing since the other thermopiles do not show it, and is most likely due to spectrally non-uniform absorption by this receiver.

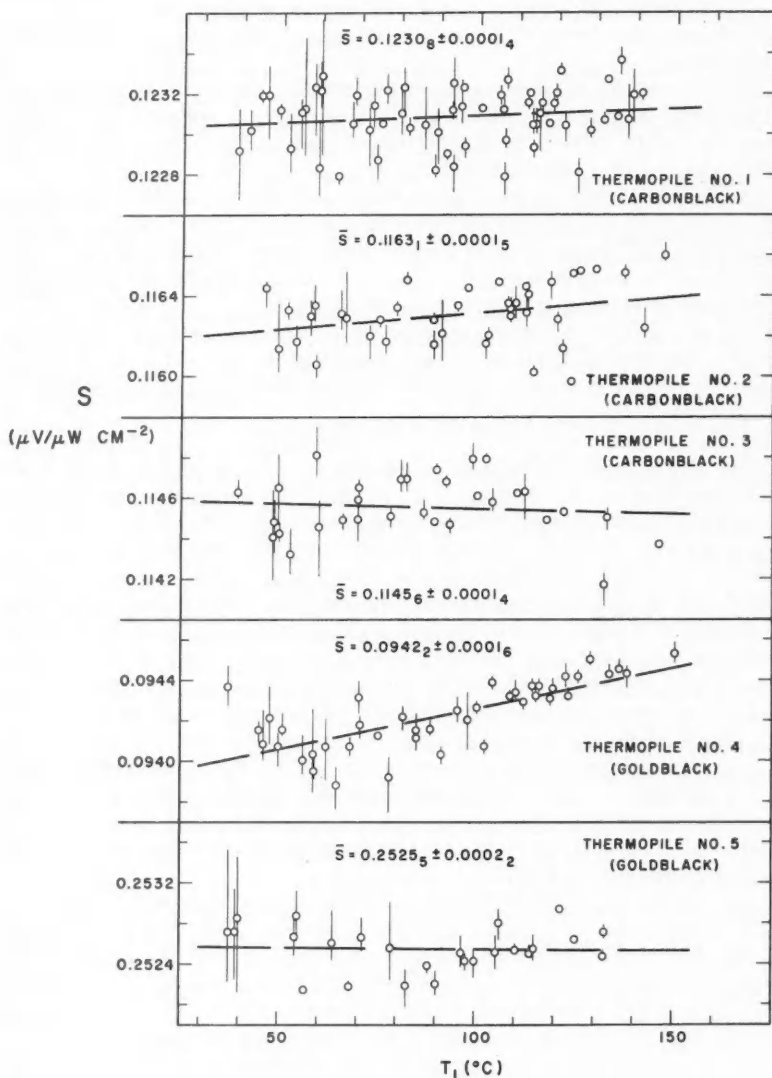


FIG. 4. Sensitivities of thermopiles. No. 1 to No. 5 as functions of the black body temperature.

Thermopile No. 4 was the one used for most of the preliminary testing, in the course of which the delicate gold-black coating was slightly damaged.

During the early stages of the calibrations an interesting feature appeared. If measurements are begun immediately after oxygen has been introduced into the system, the values of sensitivity obtained show a systematic slow decrease with time over the course of the day. Equilibrium is reached in about 24–30 hours, the total decrease during this period being a little more than 1%. The effect occurs with both gold-black and carbon-black receivers, although the latter seem to take somewhat longer to reach equilibrium. The effect also occurs if gases other than oxygen are used to displace the air. It is probably due to the slow desorption of water molecules from the blackening material, resulting in a change in the emissivity of the surface. For this reason no measurements are made until the system has had at least 48 hours' exposure to oxygen, and during the calibrations reported here the oxygen was allowed to flow continuously night and day. Leighton and Leighton (1932) reported day-to-day variations in thermopile response which seemed to be dependent on the humidity. It is not clear if the thermopiles in question were evacuated, but if not then the variations may have had the same cause.

In Fig. 5 the sensitivities of thermopiles No. 1, 2, and 4 respectively are plotted as functions of the temperature of the cold junctions,  $T_0$ . As  $T_0$  decreases,  $S$  increases, primarily because of the decrease of the thermal conductivity of the medium, and to a lesser extent because of the decrease of radiation losses. The sensitivity changes linearly with  $T_0$ , the magnitude of the change varying with the thermopile from 0.2 to 0.4% per degree, of the same order as the change of thermal conductivity with  $T_0$ . The straight lines plotted through the data are again least squares fits, and the standard deviations given are computed from the residuals of the points from the lines.

The effect of the thermal conductivity of the gas on the thermopile sensitivity is shown in Fig. 6, where the sensitivity of thermopile No. 2 is given as a function of the pressure for several gases. The sensitivity is plotted linearly in arbitrary units and the pressure on a logarithmic scale in mm Hg. There is essentially no change until the pressure is reduced to a few millimeters of mercury and then the sensitivity increases sharply until a pressure of about 1 micron is reached, where it again levels off. This is just the change one would expect as the conductivity of the gas falls from its normal value to negligible values over this same region. With oxygen the sensitivity increases by a factor of 3 when the thermopile is evacuated; with helium by a factor of 11. No difference could be detected between the responses of the thermopile in nitrogen and dry air. Careful determinations of the difference when in oxygen and nitrogen using high black body temperatures showed that the sensitivity in nitrogen is 1.0% higher than that in oxygen. The inset in Fig. 6 is a logarithmic plot of the sensitivity versus the reciprocal thermal conductivity of the gas (at atmospheric pressure). The curve shown is extrapolated towards the limiting vacuum sensitivity.

In oxygen at atmospheric pressure the time constant of the thermopile is about 2–3 seconds and so approximately 15 to 20 seconds were required for



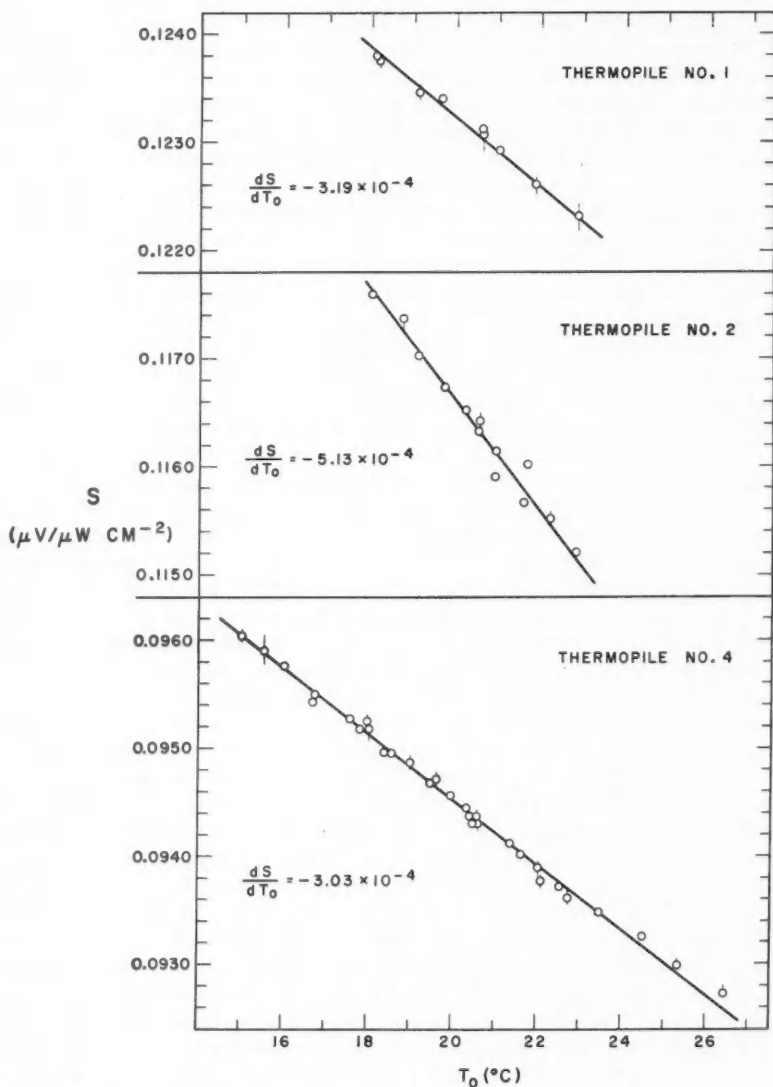


FIG. 5. Sensitivities of thermopiles No. 1, No. 2, and No. 4 as functions of the temperature of the cold junctions.



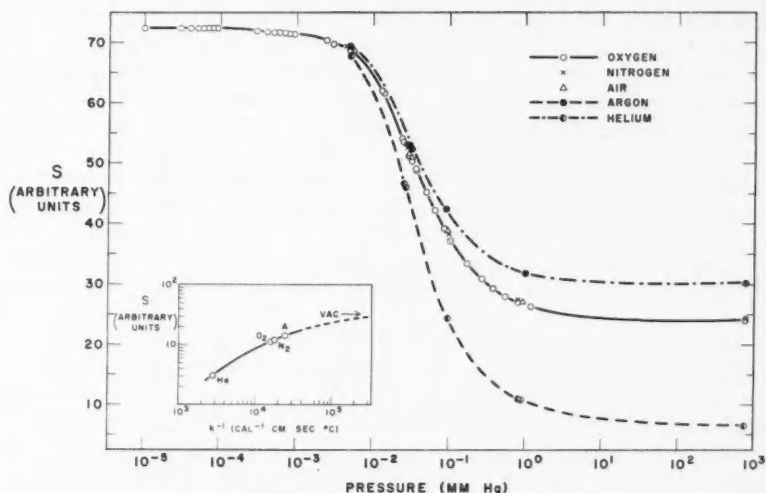


FIG. 6. Sensitivity of thermopile No. 2 as a function of the pressure. The sensitivity is plotted linearly in arbitrary units and the pressure on a logarithmic scale in mm Hg. The solid line gives results for oxygen, the broken line for argon, and the broken-dotted line for helium. Some points for nitrogen (X) and dry air ( $\Delta$ ) are also given. The inset is a logarithmic plot of the sensitivity against the reciprocal thermal conductivity of the gas (at atmospheric pressure).

the thermopile to come to equilibrium. This rather long response time is in no way objectionable in the present experiment. The time constant is dependent on the thermal conductivity of the medium in the same way as the sensitivity, and in oxygen, for example, increases by a factor of three between atmosphere and vacuum.

The amount of radiation absorbed by water vapor and carbon dioxide in the atmosphere can be roughly estimated. Measurements made at various times indicated a drop in thermopile output of 3 to 3½% when air was allowed to replace the oxygen in the system. Allowing for the 1% lower sensitivity in oxygen (but not for the "drying action" of the oxygen because the observations were made by measuring alternately in oxygen and air over short time intervals), it seems that 4 to 4½% absorption occurs. The relative humidity of the room was about 35%, and the path length, though rather indefinite, is equivalent to about 1 meter.

Figure 7 gives the results of the calibration of thermopiles 1, 2, 4, and 5 against four National Bureau of Standards radiation standards numbered C-840 to C-843 inclusive, currents of 0.250, 0.300, and 0.350 ampere being used for each lamp. The sensitivity plotted here is the ratio of the observed e.m.f. to the radiant flux density given by the lamp calibration data. The mean values obtained by calibration with the black body are also shown. The scatter in the observations decreases at the higher currents because the signal is larger and more accurately measured. The outputs obtained were in the range

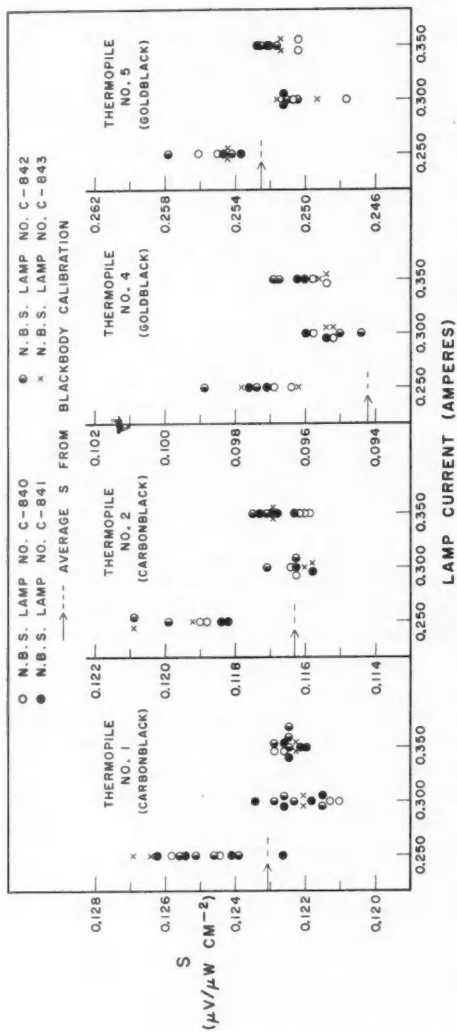


FIG. 7. Sensitivities of thermopiles No. 1, 2, 4, and 5 as obtained by calibration against four National Bureau of Standards radiation standards.

4 to 10  $\mu\text{V}$  for the eight-junction thermopiles. The data of Fig. 7 indicate that lamp orientation did not contribute appreciably to the scatter. For example, with the thermopile No. 1 and lamp C-840 the readings for 0.250 ampere are in the middle of the range, those for 0.300 ampere at the bottom, and those for 0.350 ampere at the top. If the orientation were affecting the results one would expect a correlation between the lamp and the reading. The effect of variations in lamp current is difficult to estimate, but should be small since the currents at which the lamps were calibrated were quoted to 0.001 ampere, and here current variations were not higher than 0.0003 ampere. The voltage drops across the lamps did not change appreciably during the series of measurements, but were higher than the calibration voltages quoted by 0.04% on the average.

One striking feature of the data in Fig. 7 is immediately evident—in every case the sensitivity obtained with a lamp current of 0.250 ampere is about 2% higher than that given by the other two currents, and the scatter of the points at 0.250 ampere barely overlaps the scatter for the other currents. This discrepancy is outside the tolerance of lamp calibration ( $\sim 1\%$ ). Karoli (1959) has observed the same effect with many thermopile calibrations at the Eppley Laboratory, but other workers have not mentioned such discrepancies. Guild (1937) does comment on inconsistencies of up to 1%, however. The mean sensitivity at 0.350 ampere is consistently higher than that at 0.300 ampere by about 0.4%. No reason has been found to which these variations might be attributed.

Because of these discrepancies it is rather difficult to assess the difference between the black body calibrations and the carbon lamp calibrations. For thermopiles 1, 2, and 5 the mean sensitivities obtained from the carbon lamps for currents of 0.250, 0.300, and 0.350 ampere differ on the average by +1.7%, -0.5%, and -0.2%, respectively, from those obtained from the radiation standard. For thermopile No. 4 the lamp calibrations are 3.1%, 1.1%, and 1.8% higher, respectively, and this is another indication of the non-uniform absorption by this particular receiver mentioned earlier. Since the thermopiles were used in the vertical position when calibrated with the black body and in the horizontal position when calibrated with the lamps, any systematic dependence of the sensitivities on position will appear as a difference between the two sets of results. However, it is unlikely that any significant uncertainty arises here, particularly since the distance of the receiving area of the thermopile from the front of the case was shown to be independent of position of the latter. Since the accuracy of measurement is lowest for the 0.250-ampere results, and since the instructions issued with the lamps indicate that the best results are obtained by using 0.300 to 0.350 ampere, the best estimate of the difference between the two scales of total radiation is probably -0.3%. This difference is within the tolerances quoted for the two scales.

#### CONCLUSIONS

The construction, testing, and operation of a low temperature total radiation standard have been described. It is convenient to use, simple to operate, and

is fairly adaptable so that calibrations may be performed directly with the black body or by transfer using an already calibrated thermopile. The precision of radiant energy measurements with the standard approaches the desired level of 0.1%, the standard deviations of the sets of measurements being  $\pm 0.12\%$  on the average. This figure represents the statistical uncertainties in day-to-day measurements over a rather long period. Systematic errors have been discussed, and on the basis of the results given it is felt that values of total radiation obtained with the standard are accurate to within  $\pm 0.3\%$ . The scale of total radiation defined by four National Bureau of Standards thermal radiation standards probably differs from that defined by the standard by  $-0.3\%$ .

#### ACKNOWLEDGMENTS

The author is indebted to Dr. T. M. Dauphinee for many discussions and for assistance during various phases of the work, and to Dr. H. Preston-Thomas for assistance with the manuscript. He is also grateful to Mr. J. Phillips and Mr. D. Kingswell for the extensive machining and servicing which was required during the testing of the unit.

#### REFERENCES

- BUCKLEY, H. 1934. *Phil. Mag.* **17**, 576.  
COBLENTZ, W. 1921. *Natl. Bur. Standards, Sci. Paper No.* 403.  
COBLENTZ, W. and STAIR, R. 1933. *J. Research Natl. Bur. Standards*, **11**, 79.  
COHEN, E. R., DU MOND, J. W. M., LAYTON, T. W., and ROLLETT, J. S. 1955. *Revs. Modern Phys.* **27**, 353.  
DAUPHINEE, T. M. 1953. *Can. J. Phys.* **31**, 577.  
——— 1955. *Can. J. Phys.* **33**, 275.  
DE VOS, J. C. 1954. *Physica*, **20**, 669.  
EDWARDS, D. F. 1956. University of Michigan ERI Report No. 2144-105-T.  
EPPLEY, M. and KAROLI, A. R. 1957. *J. Opt. Soc. Am.* **47**, 748.  
GOUFFÉ, A. 1945. *Revue d'Optique*, **24**, 1.  
GUILD, J. 1937. *Proc. Roy. Soc. London, A*, **161**, 1.  
JAKOB, M. 1957. *Heat transfer*, Vol. 11 (John Wiley & Sons, Inc., New York).  
KAROLI, A. R. Private communication.  
LEIGHTON, P. A. and LEIGHTON, W. G. 1932. *J. Phys. Chem.* **36**, 1882.  
VOLLMER, J. 1957. *J. Opt. Soc. Am.* **47**, 926.

# A STUDY OF AURORAL MOTIONS FROM ALL-SKY CAMERA RECORDS<sup>1</sup>

B. K. BHATTACHARYYA<sup>2</sup>

## ABSTRACT

A study of horizontal motions of visual aurorae as recorded by a 35-mm all-sky camera at Springhill (geographic 45.2° N., 75.5° W.; geomagnetic 56.5° N., 6.9° W.) near Ottawa has been carried out. The number of occurrences of motions in all the four geomagnetic directions, east, west, north, and south appears to reach its peak within a range of speed from 0 to 150 m/sec and tends to decrease with increase in speed. Very large speeds seem more frequently to be associated with motions to the west and to the south. The distribution curve of speed with the time of night appears to have two peaks, one before and another after midnight, in all the four cases. Auroral motion is predominantly westward in the early part of the night and eastward in the late hours of the night. The reversal of motion from westward to eastward direction seems to be a systematic process, the declining and inclining portions of the two curves as a function of time meeting each other somewhat before local midnight.

Auroral speeds either along or perpendicular to geomagnetic parallels of latitude increase nearly linearly with the horizontal and vertical components of the magnetic disturbance vector.

## INTRODUCTION

An intensive study of the motion of visual auroral forms has been made since it was reported by Meinel and Schulte in 1953 that auroral motion is westward in the evening, and that it reverses and becomes eastward in the morning. They based their report on the study of photographic sequence records of auroral draperies and ray structures taken at Yerkes Observatory. This report on the dependence of the direction of auroral motion on local time was not supported to the full extent by a similar study of auroral photographs taken at Saskatoon (Meek 1954). The latter study showed that westerly motions are likely to occur during the first half of the night while easterly motions are scattered throughout the night, and so a definite drift motion cannot be ascribed to all aurorae based on local time. Bless, Gartlein, and Kimball (1955) reported, from a study of visual and photographic observations of aurora at Ithaca, that when steady motions are seen in aurora they will be westward early in the night, eastward late in the night, and may be either or both ways in the middle of the night. They were also in favor of the opinion that the motions are not sharply controlled by local time. Later, Meinel (1955) stated that the relationship of the reversal of auroral motion with local time is not exact or detailed but systematic.

More recently, the study of all-sky camera photographs of aurora at Saskatoon and Flin Flon by Kim and Currie (1958) has lent support to the observations by Meek and Bless *et al.* Their results show that westward motions are

<sup>1</sup>Manuscript received July 5, 1960.

Contribution from the Division of Radio and Electrical Engineering, National Research Council, Ottawa.

Issued as N.R.C. No. 5915.

<sup>2</sup>National Research Council Postdoctorate Fellow.

characteristic of the late evening hours, eastward motions of the early morning hours, and that motions in both directions occur during a 4-hour period centered on local midnight. They did not note any relation to the local time of motions in the northward or southward direction. However, they observed an increase of speed, particularly in an east-west direction, with geomagnetic activity.

The present study is undertaken to investigate possible correlations of speeds and directions with the time of the night and with geomagnetic activity. At the same time the frequency distribution of motions in the four cardinal directions is examined with relation to the time of the night. Attention has been paid only to the motion of the displays as a whole, not to the rapid transitory motion sometimes observed within the display itself.

#### OBSERVATIONAL DETAILS

The data for the present study have been obtained from the photographs of auroral forms taken at Springhill (geographic 45.2° N., 75.5° W.; geomagnetic 56.5° N., 6.9° W.) near Ottawa during the International Geophysical Year. The records of the three months of March, April, and June, 1958, have been studied since they cover the spring equinoctial period when auroral activity was at its peak in that year. They have provided adequate data of auroral motions for this analysis.

The photographs were taken with a 35-mm all-sky camera developed by the National Research Council of Canada (Park 1957) at every 1-minute interval with alternate 'short' (4-seconds' duration) and 'long' (40-seconds' duration) exposure times. It is found that the former is too short for faint aurora whereas the latter is too long for bright aurora. Thus it has been advantageous to utilize both sets of photographs for accuracy and unambiguity.

For measurement of the motion of a chosen spot in a particular auroral display, the elevation above the horizon and the azimuth with respect to geographic north have to be determined with reasonable accuracy in two or more successive photographs. For this purpose a scale for converting the radial distance from the center of the photograph into elevation, with the help of positions of known stars, was prepared. Because of the limited accuracy of the conversion chart for points near the horizon, in general only auroral forms at an elevation greater than 10° have been studied. On the top of the screen of the projector, the above scale and another circular one for noting the azimuth have been placed.

The elevation is used to calculate the slant range of the spot with the help of the following formula, assuming a suitable height above the surface of the earth for the auroral forms:

$$R = -\gamma_0 \sin E + [(\gamma_0 + h)^2 - (\gamma_0 \cos E)^2]^{\frac{1}{2}}$$

where  $R$  is the slant range,  $\gamma_0$  the earth's radius (6370 km),  $E$  the elevation, and  $h$  the height.

The great concentration of auroral heights in the band 90–110 km has been demonstrated from a study of the distribution of heights of all the measured points scattered in the lower, middle, and upper parts of all forms of aurora

as observed in southern Norway from 1911-44 (Störmer 1955). Very great heights up to 1000 km or more have sometimes been measured, mostly in the case of auroral rays and in particular the sunlit rays. Individual studies of the height distribution of different auroral forms such as homogeneous arcs, rayed arcs, bands with rayed structure, cloudlike aurora or diffuse surfaces, draperies and pulsating surfaces, also show a strong accumulation of heights around 100 km. So it is thought legitimate to assume an auroral height of 100 km. A variation in height will induce only a random error in the calculation of speed.

The slant range of the spot, calculated by the above formula, is then converted to ground range for a flat earth. The assumption of a flat earth is justified since auroral forms close to the horizon were omitted from this study.

The co-ordinates, i.e., the ground range and azimuth of the vertical projection of a particular spot on the earth's surface, were then used to plot the spot on the special U.S.-Canada I.G.Y. auroral plotting map, which is a Murdoch's third projection taking true scale along parallels  $44^{\circ}$  N. and  $74^{\circ}$  N. of geographic latitude. On this map the geomagnetic latitude lines were superposed at intervals of 1 degree from  $50^{\circ}$  N. to  $62^{\circ}$  N. Thus the horizontal displacement of auroral forms either along or perpendicular to geomagnetic parallels of latitude could be determined. The speed was calculated by noting the time required for the displacement, 42 seconds when a 'short' exposure photograph followed a 'long' exposure one, 78 seconds in the reverse case, and 2 minutes if only either 'short' or 'long' exposures were read. The time was always taken from the center of any exposure interval.

In this study the motions of different auroral forms such as arcs, rayed arcs, rays, patches, bands, draperies, and glow have been observed and measured. Of these, arcs and rayed arcs made up 52% of the total, rays 32%, patches and bands 13%, and draperies and glows 3%. An attempt has been made, as far as practicable, to determine from successive photographs the displacements of a particular spot at the lowermost edge of the form under study. Auroral forms are found to move in all directions though the east-west component of motion is, in most cases, higher than the north-south one.

Identification of the same spot of a particular form in successive photographs is, indeed, a difficult problem and a major source of error. The accuracy of the measurements of auroral motions is greatly enhanced as the time interval used in the measurement decreases. This is because the chosen spot can be located with less ambiguity and with greater accuracy over smaller intervals of time. However, these intervals have to be increased in cases of slowly moving auroras such as homogeneous arcs.

The total time duration over which a particular spot can be identified has been found to vary from 2 minutes to more than 15 minutes in the case of rayed arcs. Homogeneous arcs are sometimes observed to retain identity for more than half an hour. Most of the forms studied, however, diffuse or change shape within 10 minutes. In most cases it has been possible to measure the motion of the same form more than once at successive intervals of 78 seconds, 42 seconds, or 120 seconds. These measurements indicate clearly that the magnitude and direction of motion do not generally remain constant for the

whole duration of the form. Each single measurement of motion has been treated as an individual case. Only in the study of the correlation of auroral speed with geomagnetic activity have the speeds been averaged over 5 minutes.

Eastern Standard Time, 5 hours less than the Greenwich time, has been used in this study. This time is the local time at the 75th meridian and hence very close to the local time of Springhill.

*Motion Along and Perpendicular to Geomagnetic Parallels of Latitude*

The whole range of speed from 0 to 1200 m/sec is divided into separate classes of width 50 m/sec. Speeds higher than 1200 m/sec are grouped in one class. The number of occurrences for different ranges of speeds along and perpendicular to the geomagnetic parallels of latitude is plotted in Fig. 1.

The number of occurrences of motions in any of the four cardinal directions appears to reach its peak within a range of speed from 0 to 150 m/sec and tends to decrease with the increase in speed. Statistical tests indicate that the number of occurrences reaches its peak within the range of 50 to 100 m/sec for eastward and northward motions and within 100 to 150 m/sec for westward and southward motions. It is also found that the number of occurrences is greater than half the maximum number observed when the speed is less than 300 m/sec for eastward motion, 250 m/sec for westward motion, 150 m/sec for northward motion, and 300 m/sec for southward motion. The median values of speed are 170, 200, 110, and 160 m/sec for auroras moving to the east, west, north, and south respectively.

For speeds greater than 700 m/sec, 20 cases of westward, 14 cases of southward, 5 cases of eastward, and 1 case of northward motions have been noted (Fig. 1). It seems, therefore, that the large speeds are more frequently associated with westward and southward motions than with eastward and northward ones.

To determine the distribution of the direction of motion in relation to a particular hour of the night, the number of occurrences for each direction separately has been plotted against Eastern Standard Time in Fig. 2. From this plot it is evident that motion is predominantly westward before local midnight and eastward after local midnight. The number of westward motions is maximum between 21 and 22 hours and systematically decreases after that. In the case of eastward motion the number begins to increase from about 22 hours, reaches its peak between 01 and 02 hours, and then begins to decrease rather sharply. It seems that the reversal of motion from geomagnetic west to east is a systematic process, the declining and inclining portions of the two curves meeting each other somewhat before local midnight.

Southward-moving auroral forms outnumber those in the reverse direction up to 02 hours. The peak density of south motions occurs between 01 and 02 hours. The number of northward-moving forms begins to increase after 23 hours, reaches its peak between 23 and 00 hours, and thereafter begins to decrease systematically. From 02 hours on, the number of north-going forms is greater than those moving to the south.

The variation of speed with the hour of the night is shown in Fig. 3, where the average speed over all the values obtained in the specific interval of 1 hour is taken. The figure indicates that there is a tendency for two peaks in



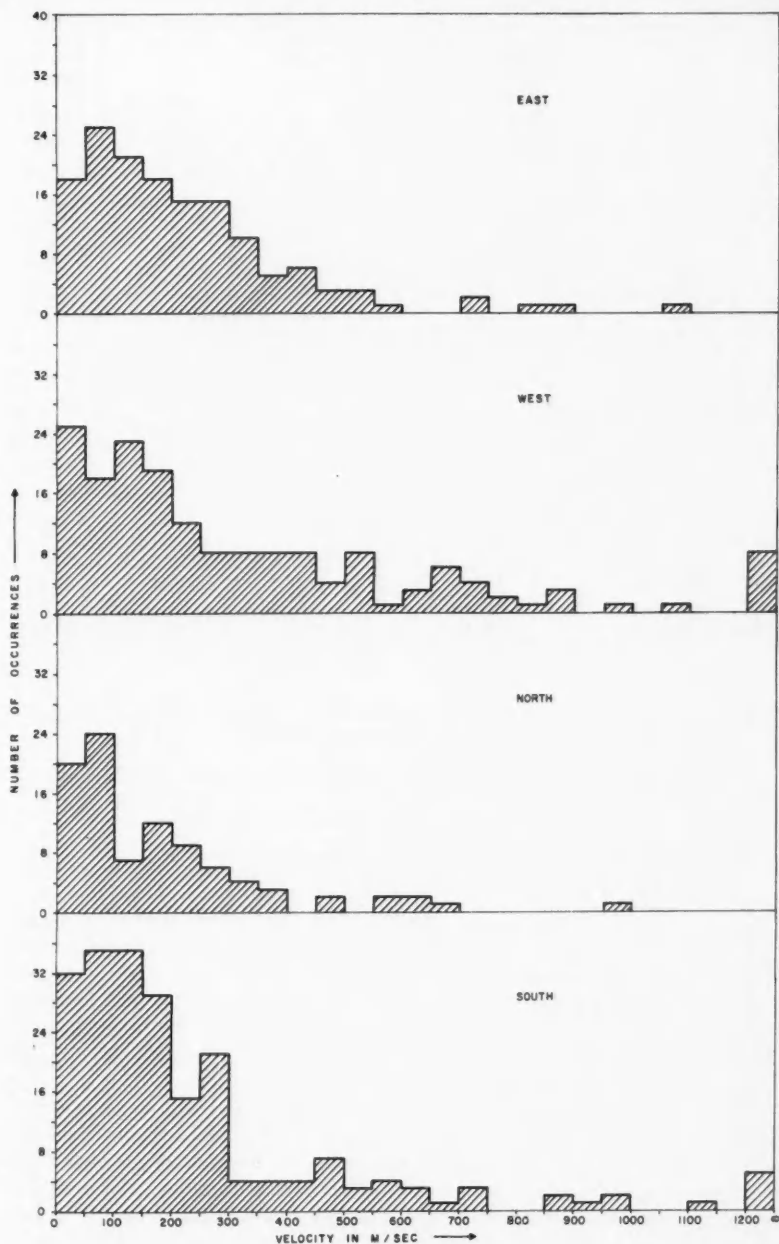


FIG. 1. Histogram of the number of occurrences of auroral motion along and perpendicular to a geomagnetic parallel of latitude as a function of 50 m/sec ranges of speed of auroras.

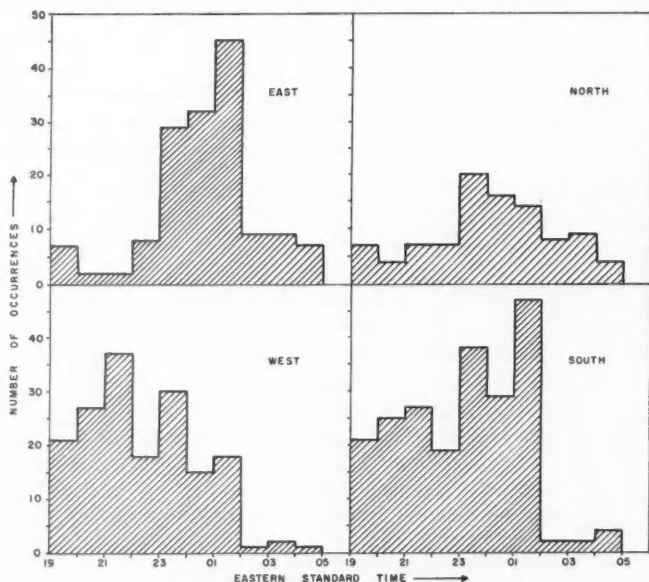


FIG. 2. Histogram of the number of occurrences of auroral motion along and perpendicular to a geomagnetic parallel of latitude as a function of Eastern Standard Time.

each of the four distributions. The approximate time-intervals during which these peaks occur, are shown in the following table.

Direction of motion	Approximate time intervals for the occurrence of maximum speeds
East	22:00-23:00 hr and 02:00-03:00 hr
West	21:00-22:00 hr and 02:00-03:00 hr
North	20:00-21:00 hr and 02:00-03:00 hr
South	19:00-20:00 hr and 03:00-04:00 hr

#### *Auroral Speed and Geomagnetic Activity*

Auroral motions are not observed on every night when aurora is visible at Springhill. Even on those nights when motions are abundant, they are not continuous throughout the night. They can usually be detected only in a small fraction of the total nighttime duration.

This imposes a restriction on the choice of a suitable index of geomagnetic activity to which auroral speeds may be related. The *K*-index, so often used, seems to be unsuitable for our purpose, since it is measured on a 3-hour basis. The necessity of an index at intervals of smaller time ranges has been met by measuring the deviations of the horizontal and vertical magnetic field components at intervals of 5 minutes. These intervals are again selected in

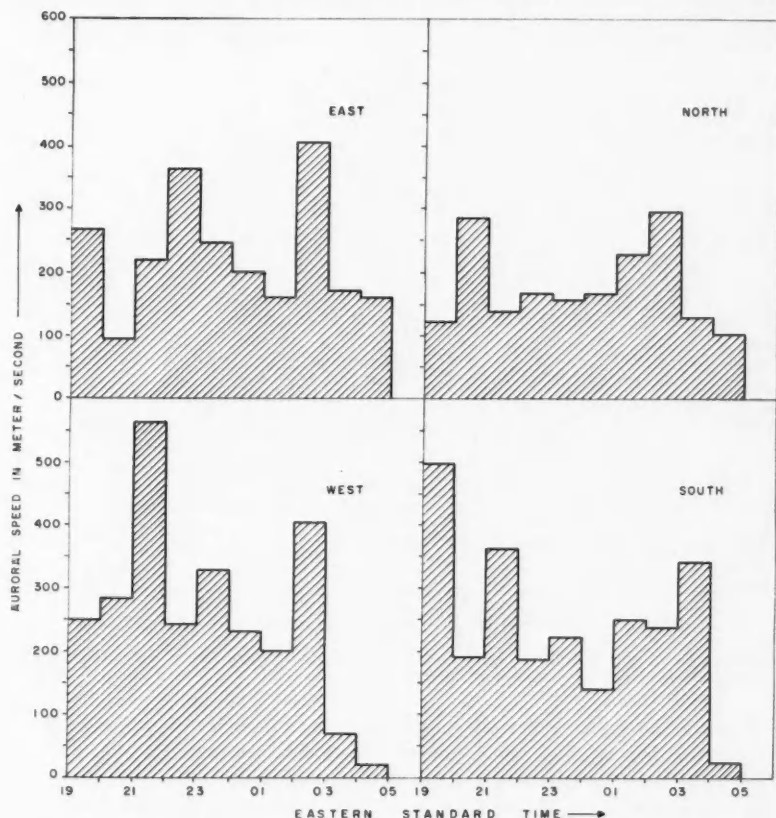


FIG. 3. Histogram of auroral speed along and perpendicular to a geomagnetic parallel of latitude versus time.

accordance with the times during which motions are observed. The geomagnetic data were taken from the magnetograms of Agincourt (geographic  $43.8^{\circ}$  N.,  $79.3^{\circ}$  W.; geomagnetic  $55^{\circ}$  N.,  $11.6^{\circ}$  W.).

In this case the auroral speeds, where possible, are averaged over a 5-minute interval. Average auroral speeds above 600 m/sec are discarded from this study because of the abrupt decrease in the number of occurrences and the total range of speed from 0 to 600 m/sec is divided into six equal intervals. The deviations in the horizontal and vertical magnetic field components corresponding to each auroral speed interval averaged without regard to the time of the night or the month. These average values are denoted by  $\Delta H_1$  and  $\Delta Z_1$  for the horizontal and vertical components respectively. The variations of  $\Delta H_1$  and  $\Delta Z_1$  with auroral speed are plotted in Figs. 4 and 5 for east-west and north-south movements respectively.

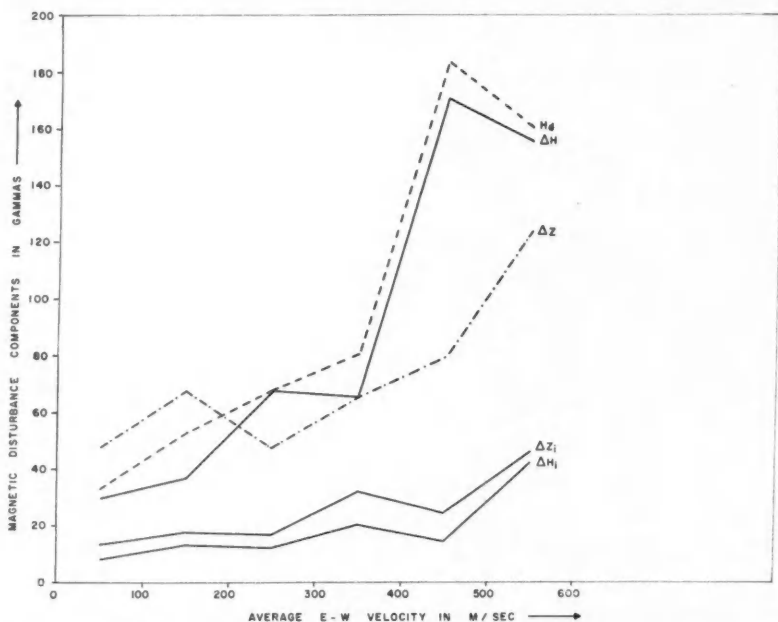


FIG. 4. Plot of the different components of geomagnetic disturbance vectors and indices versus 100 m/sec ranges of east-west auroral speed.

Besides the above index of geomagnetic activity, the magnitudes of the horizontal and vertical components of magnetic disturbance vectors have been calculated at intervals of 5 minutes on the basis of the monthly quiet-day average values in March, April, and June, 1958. Let  $H$ ,  $D$ , and  $Z$  represent the 5-minute average values of the horizontal component, declination, and vertical component of the magnetic field respectively. If  $H_q$ ,  $D_q$ , and  $Z_q$  be the monthly quiet-day average values, then

$$\Delta H = H - H_q \text{ (positive northward),}$$

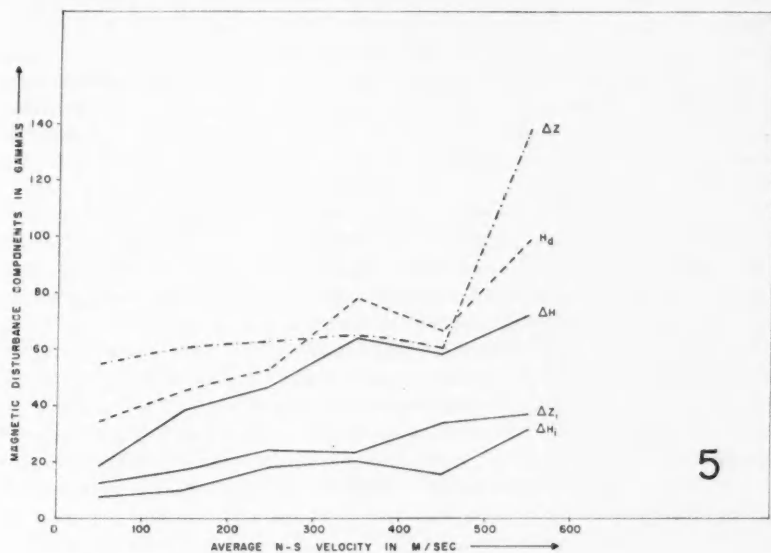
$$\Delta D = D - D_q \text{ (positive eastward),}$$

$$\Delta Z = Z - Z_q \text{ (positive vertically downward),}$$

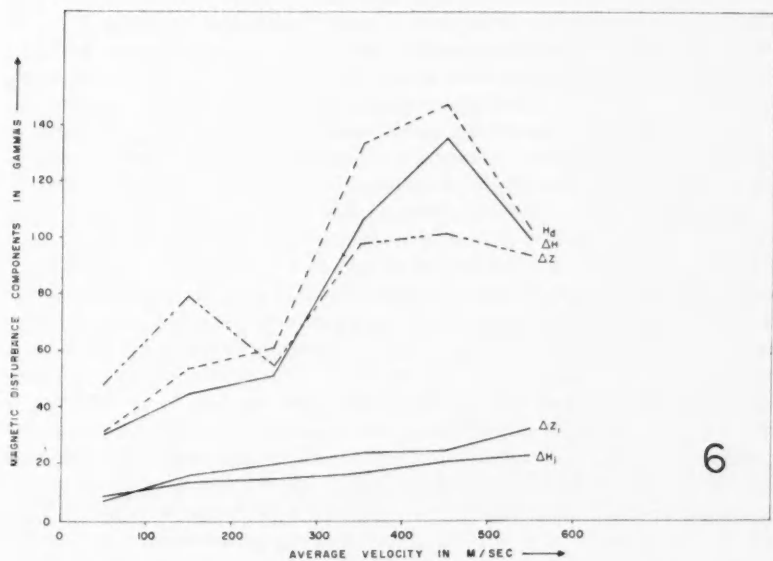
are the disturbance values.  $\Delta H$  and  $\Delta D$  are used to determine  $H_d$ , the magnitude of the horizontal disturbance vector.

The average values of the magnitudes of  $\Delta H$ ,  $H_d$ , and  $\Delta Z$  are then calculated for each of the six ranges of auroral speed. The curves of the average values of  $\Delta H$ ,  $H_d$ , and  $\Delta Z$  are plotted as a function of auroral speed in Figs. 4 and 5.

A study of the above figures shows approximately a linear relationship between geomagnetic activity and auroral speed either along or perpendicular to the geomagnetic parallels of latitude. All the curves in Figs. 4 and 5 show



5



6

FIG. 5. Plot of the different components of geomagnetic disturbance vectors and indices versus 100 m/sec ranges of north-south auroral speed.

FIG. 6. Plot of the different components of geomagnetic disturbance vectors and indices versus 100 m/sec ranges of average speed of aurora.

the same general tendency. In the case of east-west motion, the horizontal component seems to be more affected than the vertical component, while this is reversed when we consider north-south motion.

In Fig. 6 is presented the variation of all the above magnetic field components with the 5-minute averages of total auroral speed irrespective of the direction of motion. The characteristics of the curves are similar to those in Figs. 4 and 5.

#### DISCUSSION

Studies of auroral motions determined from radio data have been carried out by various workers (Bowles 1954; Bullough and Kaiser 1955; Nichols 1957). Of these, we shall quote the recent results by Nichols obtained by a study of frequency spectra of auroral radio echoes from east and west of geomagnetic north at College, Alaska (geographic  $64.9^\circ$  N.,  $147.8^\circ$  W.; geomagnetic  $64.7^\circ$  N.,  $256.5^\circ$  E.) at two frequencies 106 Mc/s and 41.15 Mc/s. These results show that the motions are generally horizontal and in the geomagnetic east-west directions, though north-south components are present. The technique, adopted by Nichols, unfortunately did not allow the north-south component of motion to be accurately measured. Nichols also noted the preponderance of auroral motion in the westward sense in the early part of the night and in the eastward sense just before and after midnight.

The results obtained from the present study are mostly in accord with those of radio data, so far as the local time distribution of the horizontal east-west motion is concerned. It is to be noted, however, that the motion as detected by radar cannot be associated with any particular auroral form with certainty. From the photographic auroral data there is an obvious advantage of measuring the north-south component of motion with sufficient accuracy. This component has been observed to be present at all hours of the night for practically all auroral forms. The number of southward motions is greater in the early part of the night and less in the late hours of the night than that of northward motions.

The measurements of *E*-layer wind drifts (Briggs and Spencer 1954) indicate a strong semidiurnal variation which in the northern hemisphere is a wind to the south at 21 and 09 hours and to the north at 15 and 03 hours local time. Chapman (1953) also found similar results in his measurements at Ottawa, though his data was mostly confined to the daytime. These northward and southward movements of a conducting medium across the lines of force of earth's magnetic field will produce an electromotive force directed toward the east and west respectively. If this electromotive force were responsible for the motion of auroral forms, we can reasonably expect peaks in the westward and eastward motions around 21 and 03 hours respectively. We note, indeed, maxima in speeds towards the west and the east between 21 and 22 hours and 02 and 03 hours respectively. This has also been observed by Lyon and Kavadas (1958) from their study of horizontal motions in radar auroral echoes.

Chapman has reported an increase of *E*-layer drift velocity with the geomagnetic *K*-indices. In this study speed of auroral forms has also been found to increase almost linearly with geomagnetic activity, the index of which may be either the horizontal or the vertical component of the magnetic disturbance

vector. Though some of the statistical properties of the *E*-layer wind drifts and motions of auroral forms tally with each other very well, they, however, move with vastly different velocities in general.

Exceedingly high velocities of auroral forms greater than 500 m/sec are often observed. The mechanism which is responsible for these velocities is not yet known. It is hard to account for these, if we consider an auroral form to be an ionized irregularity immersed in the *E*-region ionization of the ionosphere.

Recent theoretical studies of the layer drift, i.e., the drift of the ionization surrounding an irregularity in the *E*-region (Maeda 1959*a*) have shown that the drift velocity of the layer is negligible compared to the air velocity, and consequently the layer moves almost with the same velocity as that of air under normal quiet conditions. When the electron (or ion) density of an irregularity exceeds that of the surrounding ionization, the drift velocity of the irregularity is always smaller than that of the medium (Weekes 1957; Kato 1959). So if it is assumed that the electron (or ion) density of auroral forms is greater than that of the surrounding ionization, the auroral drift is less than the layer drift, which is again very small compared to the air velocity. Hence the exceedingly high velocities of the auroral form and the layer cannot be related to the general wind motion in the ionosphere. They may result from a great enhancement of the applied electric field and conductivity under magnetically disturbed conditions. Maeda (1959*b*) has reported the existence of strong polar fields (about  $5 \times 10^4$  e.m.u.) which are responsible for  $S_d$  field variation. Such strong polar fields result in a layer drift velocity which may be more than 100 times that under normal conditions (Kato 1959). A quantitative estimate of the enhancement of the electric field under disturbance conditions may prove to be valuable in solving the problem of the generation mechanism of the exceptionally high velocities of some visual auroral forms.

If the motion of auroral forms as we observe them is due to that of the beam of ionizing particles which cause the aurora, then the above picture is not true. These particles will move independently of the drift of the ionization of the layer on which they are incident. We may, then, picture the motion of visible aurora as the appearance of the same ionizing agent at different places at successive instants of time. In this picture we have, however, implicitly assumed that the luminosity of the auroral form at a certain place decays very quickly due to various factors such as de-excitation and diffusion.

#### ACKNOWLEDGMENTS

The author is deeply indebted to Dr. A. G. McNamara and Dr. P. M. Millman for many helpful discussions and suggestions. The assistance of Mr. F. R. Park and Mr. F. MacDonald in the collection and analysis of data is gratefully acknowledged.

#### REFERENCES

- BLESS, R. C., GARTLEIN, C. W., and KIMBALL, D. S. 1955. *Astrophys. J.* **122**, 205.  
BOWLES, K. L. 1954. *J. Geophys. Research*, **59**, 553.  
BRIGGS, B. H. and SPENCER, M. 1954. *Repts. Progr. in Phys.* **17**, 245.  
BULLOUGH, K. and KAISER, T. R. 1955. *J. Atmospheric and Terrest. Phys.* **6**, 198.  
CHAPMAN, J. H. 1953. *Can. J. Phys.* **31**, 120.  
KATO, S. 1959. *Rept. Ionosphere Space Research Japan*, **13**, 62.

- KIM, J. S. and CURRIE, B. W. 1958. *Can. J. Phys.* **36**, 160.  
LYON, G. F. and KAVADAS, A. 1958. *Can. J. Phys.* **36**, 1661.  
MAEDA, H. 1959a. Rept. Ionosphere Space Research Japan, **13**, 79.  
——— 1959b. *J. Geomagnetism and Geoelec.* **10**, 66.  
MEEK, J. H. 1954. *Astrophys. J.* **120**, 602.  
MEINEL, A. B. 1955. *Astrophys. J.* **122**, 206.  
MEINEL, A. B. and SCHULTE, D. H. 1953. *Astrophys. J.* **117**, 454.  
NICHOLS, B. 1957. *J. Atmospheric and Terrest. Phys.* **11**, 292.  
——— 1957. Sci. Rept. No. 1, Geophysical Institute, University of Alaska.  
PARK, F. R. 1957. *Bull. Radio and Elec. Eng. Div., National Research Council, Ottawa*, **7**, 1.  
STÖRMER, C. 1955. *The polar aurora* (Oxford University Press, London).  
WEEKES, K. 1957. *J. Atmospheric and Terrest. Phys. Special Supplement, Polar Atmosphere Symposium, Part II*, 12.



# THE RAMAN SPECTRUM OF METHANE<sup>1</sup>

M. A. THOMAS<sup>2</sup> AND H. L. WELSH

## ABSTRACT

The Raman spectrum of CH<sub>4</sub> was obtained with a spectral resolution of  $\sim 0.3$  cm<sup>-1</sup>, and rotational analyses of the  $\nu_2$  and  $\nu_3$  bands were carried out. The  $B_0$  values obtained from the  $\nu_2$  and  $\nu_3$  bands are  $5.240 \pm 0.002$  and  $5.2406 \pm 0.0011$  cm<sup>-1</sup>, respectively; the value of  $r_0$  determined from the latter is  $1.09403 \pm 0.00016$  Å. The rotational levels of the  $v = 1$  state of  $\nu_2$  are double with  $B_1$  values of 5.313 and 5.379 cm<sup>-1</sup>. The rotational levels of the  $v = 1$  state of  $\nu_3$  do not follow the theoretical formulae exactly; the deviations for the 13 branches observed can be expressed by using three different  $B_1$  values, 5.178, 5.195, and 5.212 cm<sup>-1</sup>. The value of  $\zeta_3$  is 0.054. The band origins are:  $\nu_1 = 2916.7$ ,  $\nu_2 = 1533.6$ , and  $\nu_3 = 3018.9$  cm<sup>-1</sup>.

## INTRODUCTION

The present communication is the first in a series in which it is intended to present high resolution data on the Raman spectra of methane and its deuterated species in the gaseous state. The Raman spectra of these molecules are of interest from many points of view, among which the following might be mentioned:

(a) The rotational constant,  $B_0$ , must be obtained from vibrational spectra since, at least for CH<sub>4</sub> and CD<sub>4</sub>, there is no pure rotational spectrum. Very weak rotational Raman spectra should exist for CH<sub>3</sub>D, CH<sub>2</sub>D<sub>2</sub>, and CHD<sub>3</sub>, but these have not yet been observed.

(b) Accurate determination of the moments of inertia of these molecules can lead to precise data on the effect of isotopic substitution on the geometry of the ground state of the methane molecule.

(c) All of the fundamental vibrations of all the molecules in the series are active in the Raman effect. For the degenerate bands of the spherical and the symmetric tops the rotational selection rules are not the same in Raman scattering and in infrared absorption. In particular, the rotational constant,  $A$ , of the symmetric tops can be obtained in principle from the rotational analysis of any one degenerate Raman band.

(d) Analyses of infrared and Raman spectra of the methanes which have been made up to the present show the existence of divergences from the simple theory. With high resolution data available for both types of spectra, these divergences will be more clearly delineated.

Ordinary methane (point-group  $T_d$ ) has four fundamental vibrations: one totally symmetric,  $\nu_1(a_1)$ , one doubly degenerate,  $\nu_2(e)$ , and two triply degenerate,  $\nu_3, \nu_4(f_2)$ . All the fundamentals are active in the Raman effect; in infrared absorption  $\nu_3$  and  $\nu_4$  are active from symmetry reasons and  $\nu_2$  is weakly active as a consequence of Coriolis interaction with  $\nu_4$  (Burgess, Bell, and Nielsen

<sup>1</sup>Manuscript received July 6, 1960.

Contribution from the McLennan Laboratory, University of Toronto, Toronto, Ontario. This research was supported in part by funds from the National Research Council of Canada.

<sup>2</sup>Holder of the Vincent Massey Scholarship, 1955-56.

1953). The rotational structure of the  $\nu_3$  Raman band was photographed by Stoicheff, Cumming, St. John, and Welsh (1952) with a reciprocal linear dispersion of  $27 \text{ cm}^{-1}$  per mm using Hg 2537 as the exciting line. An analysis of the 68 observed maxima showed unexpected irregularities in the rotational constants of the upper vibrational state. The  $\nu_2$  band, studied at a dispersion of  $10.5 \text{ cm}^{-1}$  per mm in the visible region by Feldman, Romanko, and Welsh (1955), showed a splitting of the rotational levels of the upper vibrational state into two components. The  $\nu_2$  and  $\nu_3$  bands of  $\text{CD}_4$ , investigated at the same dispersion by Shepherd and Welsh (1957), showed irregularities very similar to those observed in the spectrum of  $\text{CH}_4$ .

In the present investigation the Raman spectrum of methane was photographed with a high light-power grating instrument which gave a reciprocal linear dispersion of  $6 \text{ cm}^{-1}$  per mm at  $4358 \text{ \AA}$  and a practical resolving power of  $\sim 0.3 \text{ cm}^{-1}$ . The Raman source was a Pyrex tube, 7.3 m long and 15 cm in diameter, illuminated by four high-current externally cooled mercury arcs, each about 3 m long. A multiple reflection mirror system (Welsh, Stansbury, Romanko, and Feldman 1955) was incorporated in the Raman tube.

#### THE $\nu_1$ BAND

The totally symmetric CH stretching frequency,  $\nu_1$ , was observed as a strong unresolved Q branch with a maximum at  $2917.0 \pm 0.1 \text{ cm}^{-1}$ . This value is the average of four measurements of the shift from Hg 4047 and Hg 4078 on two spectrograms. The band is degraded slightly towards higher frequency shifts and has a half-width of  $\sim 0.9 \text{ cm}^{-1}$ , indicating that  $B_1 - B_0$  is positive and very small. The origin of the band,  $\nu_0$ , is probably at  $2916.7 \text{ cm}^{-1}$ , the frequency of the sharp low-frequency edge of the band. A weak line at  $3065.8 \text{ cm}^{-1}$ , which could not be assigned to the rotational structure of the  $\nu_3$  band, is probably  $2\nu_2$  in Fermi resonance with  $\nu_1$ .

#### THE $\nu_2$ BAND

A microphotometer trace of the  $\nu_2$  band is shown in Fig. 1, and the frequency shifts averaged from the measurement of two spectrograms with exposure times of  $\sim 50$  hours are given in Table I. The measured frequency shifts extend from  $1345$  to  $1817 \text{ cm}^{-1}$ ; since the  $\nu_3$  band excited by Hg 4047 overlaps a part of this region, a filter solution of sodium nitrite was used to eliminate the violet Hg lines from the exciting radiation.

According to the selection rules,  $\Delta J = 0, \pm 1, \pm 2$ , the band should consist of O, P, Q, R, and S branches. The earlier analysis by Feldman, Romanko, and Welsh (1955) indicated that the rotational levels of the  $v = 1$  vibrational state are split into two sets of sublevels corresponding to two different values of the rotational constant,  $B_1$ ; following the notation of these authors, we designate the upper and lower components of the rotational levels by  $\alpha$  and  $\beta$ , respectively. A preliminary analysis showed that most of the lines observed could be assigned to  $O^\beta$ ,  $S^\beta$ ,  $Q^\alpha$ ,  $P^\alpha$ , and  $R^\alpha$  branches in agreement with the earlier work; a few weak lines observed for the first time in the present study could be identified as a  $P^\beta$  branch. The assignments are given in Table I.

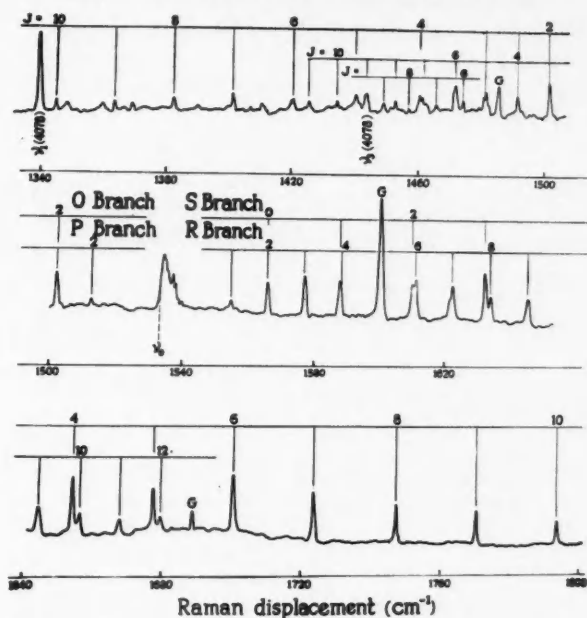


FIG. 1. Microphotometer trace of the  $\nu_2$  Raman band of methane, excited by Hg 4358. The lines marked G are grating ghosts.

TABLE I  
Raman frequencies of the  $\nu_2$  band of  $\text{CH}_4$

$\Delta\nu(\text{cm}^{-1})$	Relative intensity	Assignment	$\Delta\nu(\text{cm}^{-1})$	Relative intensity	Assignment
1344.9	1	$O^{\beta}(10)$	1536.0	6	$Q^{\alpha}(6)$
1363.7	1	$O^{\beta}(9)$	1537.4	5	$Q^{\alpha}(7)$
1382.6	1	$O^{\beta}(8)$	1538.3	4	$Q^{\alpha}(8)$
1401.7	3	$O^{\beta}(7)$	1539.2	3	$Q^{\alpha}(9)$
1420.4	2	—	1540.1	3	$Q^{\alpha}(10)$
1421.1	2	$O^{\beta}(6)$	1555.1	0	$R^{\alpha}(1)$
1425.9	0	$P^{\alpha}(11)$	1565.9	4	$S^{\beta}(0); R^{\alpha}(2)$
1433.7	0	$P^{\beta}(11)$	1577.2	5	$R^{\alpha}(3)$
1435.0	0	$P^{\alpha}(10)$	1587.7	5	$S^{\beta}(1)$
1441.1	2	$O^{\beta}(5)$	1588.3	4	$R^{\alpha}(4)$
1443.8	3	$P^{\alpha}(9); \nu_3(4078)$	1609.7	5	$S^{\beta}(2)$
1449.2	0	$P^{\beta}(9)$	1610.8	5	$R^{\alpha}(6)$
1453.1	1	$P^{\alpha}(8)$	1622.2	5	$R^{\alpha}(7)$
1457.4	0	$P^{\beta}(8)$	1631.9	7	$S^{\beta}(3)$
1461.1	2	$O^{\beta}(4)$	1633.4	3	$R^{\alpha}(8)$
1461.9	1	—	1645.0	3	$R^{\alpha}(9)$
1463.2	0	$P^{\alpha}(7)$	1654.5	8	$S^{\beta}(4)$
1466.0	0	$P^{\beta}(7)$	1656.4	2	$R^{\alpha}(10)$
1472.0	2	—	1667.8	1	$R^{\alpha}(11)$
1472.6	3	$P^{\alpha}(6)$	1677.2	7	$S^{\beta}(5)$
1474.8	1	$P^{\beta}(6)$	1679.2	0	$R^{\alpha}(12)$
1481.4	2	$O^{\beta}(3)$	1700.0	8	$S^{\beta}(6)$
1482.3	4	$P^{\alpha}(5)$	1723.2	7	$S^{\beta}(7)$
1492.3	2	$P^{\alpha}(4)$	1746.6	5	$S^{\beta}(8)$
1502.3	4	$O^{\beta}(2); P^{\alpha}(3)$	1770.1	4	$S^{\beta}(9)$
1512.8	0	$P^{\alpha}(2)$	1793.5	3	$S^{\beta}(10)$
1534.7	10	$Q^{\alpha}_{\text{max}}$	1817.1	1	$S^{\beta}(11)$

Since the  $O^{\beta}$  and  $S^{\beta}$  lines form the best-defined series and extend over the greatest frequency range, they were used for the determination of the ground state constants. The frequency shifts expected from the theory can be represented by the equation

$$(1) \quad \Delta\nu = a + bm + cm^2 - dm^3 - em^4$$

where  $m = 2J+3$ ,  $J = 0, 1, 2, \dots$  for the  $S$  branch and  $m = -2J+1$ ,  $J = 2, 3, 4, \dots$  for the  $O$  branch, and the coefficients are the following functions of the rotational constants,  $B_v$ , and the centrifugal stretching constants,  $D_v$ :

$$(2) \quad \begin{aligned} a &= \nu_0 + (3/4)(B_1 - B_0) - (9/16)(D_1 - D_0), \\ b &= (B_1 + B_0) - (3/2)(D_1 + D_0), \\ c &= (1/4)(E_1 - B_0) - (11/8)(D_1 - D_0), \\ d &= (1/2)(D_1 + D_0), \\ e &= (1/16)(D_1 - D_0). \end{aligned}$$

The five coefficients in eq. (1) were determined by a least squares method from 20 observed  $O$  and  $S$  frequencies with the aid of an IBM 650 electronic computer. The values obtained along with the probable errors are:

$$\begin{aligned} a &= 1533.69 \pm 0.03, \\ b &= 10.619 \pm 0.003, \\ c &= 0.0347 \pm 0.0003, \\ d &= (1.38 \pm 0.09) \times 10^{-4}, \\ e &= (0.040 \pm 0.006) \times 10^{-4}. \end{aligned}$$

The molecular constants thus determined are:

$$\begin{aligned} \nu_0 &= 1533.6 \text{ cm}^{-1}, \\ B_0 &= 5.240 \pm 0.002 \text{ cm}^{-1}, \\ B_1^{\beta} &= 5.379 \pm 0.002 \text{ cm}^{-1}, \\ D_0 &= (1.1 \pm 0.2) \times 10^{-4} \text{ cm}^{-1}, \\ D_1^{\beta} &= (1.7 \pm 0.2) \times 10^{-4} \text{ cm}^{-1}. \end{aligned}$$

The  $B$  values obtained here differ considerably from the values,  $B_0 = 5.270$  and  $B_1^{\beta} = 5.392 \text{ cm}^{-1}$ , determined by Feldman *et al.* (1955). The differences seem to be due to inaccurate frequency measurements in the earlier work, particularly for the higher members of the  $O$  series which were very weak on the spectrograms.

Since the values of  $\nu_0$ ,  $B_0$ , and  $D_0$  obtained from this analysis could be assumed to be reasonably accurate they were used in determining the upper state constants for the  $P$  and  $R$  branches. The rotational terms of the upper state,  $F_1(J')$ , can be calculated from the relation,

$$(3) \quad F_1(J') = \nu - \nu_0 + F_0(J''),$$

where  $\nu$  is the observed frequency of the rotational line in question and  $F_0(J'')$  is the corresponding ground state term calculated from  $B_0$  and  $D_0$ . The upper state constants,  $B_1$  and  $D_1$ , of a given branch can then be determined from the equation,

$$(4) \quad F_1(J)/J(J+1) = B_1 - D_1J(J+1),$$

by plotting the left-hand side against  $J(J+1)$  as shown in Fig. 2. The points for the  $P^\alpha$  and  $R^\alpha$  lines fall on a single straight line the equation of which,

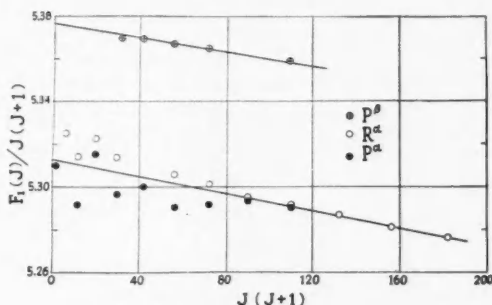


FIG. 2. Analysis of the  $\nu_2$  Raman band of methane.

determined by a least squares computation, gave  $B_1^\alpha = 5.313$  and  $D_1^\alpha = 2 \times 10^{-4} \text{ cm}^{-1}$ . The lines of the weaker  $P$  branch fall on a different straight line which gives a  $B_1$  value very close to  $5.379 \text{ cm}^{-1}$ ; these lines should therefore be identified as a  $P^\beta$  branch. If an  $R^\beta$  branch is present in the Raman spectrum it is too weak to be observed.

The present study of the  $\nu_2$  band has thus confirmed the main features of the analysis of Feldman *et al.* (1955) and has definitely improved the accuracy of the rotational constants. When the constants determined from the analysis were used to calculate the Raman frequencies, the average deviation from the observed frequencies was  $0.05 \text{ cm}^{-1}$  for the  $O^\beta$  and  $S^\beta$  lines,  $0.1 \text{ cm}^{-1}$  for the  $P^\alpha$  and  $R^\alpha$  lines, and  $0.2 \text{ cm}^{-1}$  for the  $P^\beta$  lines.

A weak infrared activity of the  $\nu_2$  vibration was observed by Burgess, Bell, and Nielsen (1953), who measured the frequencies of 83 lines in the band. Many of these coincide with observed Raman frequencies, and, with the constants obtained from the Raman spectrum, some 68 of the infrared lines can be identified as members of  $P^\alpha$ ,  $P^\beta$ ,  $Q^\alpha$ ,  $Q^\beta$ ,  $R^\alpha$ , and  $R^\beta$  branches. The remaining infrared lines appear to result from perturbations of individual rotational levels. Rotational perturbations probably account also for three Raman frequencies (Table I) which could not be assigned in the above analysis.

#### THE $\nu_3$ BAND

As was shown first by Teller (1934), the rotational levels of the upper vibrational state ( $F_2$  species) of the  $\nu_3$  band is split into three sublevels as a

consequence of the Coriolis interaction of the rotational and vibrational motions. The energies of the three rotational sublevels, designated +, 0, and -, are given by:

$$\begin{aligned} F_1^+(J) &= B_1J(J+1) + 2B_1\zeta_3(J+1), \\ (5) \quad F_1^0(J) &= B_1J(J+1), \\ F_1^-(J) &= B_1J(J+1) - 2B_1\zeta_3(J), \end{aligned}$$

where  $\zeta_3$  is the vibrational angular momentum of the  $v = 1$  state of the  $\nu_3$  vibration and the centrifugal stretching terms are omitted. The selection rules for the 1-0 vibrational Raman transition are  $\Delta J = 0, \pm 1, \pm 2, J' + J'' \geq 2$ , and transitions are allowed to all three sublevels of the upper state; the Raman spectrum therefore consists of 15 branches. The relative intensities of the different branches in the region of high  $J$  values, as given by Teller (1934), are:

$$\begin{aligned} (6) \quad S^+ : S^0 : S^- &= O^- : O^0 : O^+ = 15 : 5 : 1, \\ R^+ : R^0 : R^- &= P^- : P^0 : P^+ = 10 : 9 : 3, \\ Q^+ : Q^0 : Q^- &= 6 : 9 : 6. \end{aligned}$$

In the earlier work at rather low dispersion (Stoicheff *et al.* 1952) evidence for most of the branches was found; however, the frequency analysis gave different values of  $B_1$  for various branches.

The  $\nu_3$  Raman band as photographed in the present investigation extends from  $2700 \text{ cm}^{-1}$  to  $3300 \text{ cm}^{-1}$ ; this region also includes the  $\nu_1$  and  $2\nu_2$  bands as single  $Q$  branch maxima. Spectrograms of the band excited by Hg 4047 are reproduced in Fig. 3. The tilting of the spectral lines which is noticeable in the spectrograms is due to the optical arrangement of the high light-power grating spectrograph; the tilting, which increases with the distance along the spectrum from the optic axis, is an inconvenience but does not introduce appreciable error in the frequency measurements since the iron comparison lines are also tilted. A microphotometer trace of the same spectrograms is shown in Fig. 4. The higher frequency shifts in the  $\nu_3$  band excited by Hg 4047 are overlapped to some extent by the  $\nu_2$  band excited by Hg 4358; the  $\nu_3$  band was therefore also photographed with Hg 4358 as the exciting line and a part of this spectrogram showing the region of higher frequency shifts is reproduced in Fig. 3(c). The measured Raman frequencies are given in Table II, the numbering of the lines corresponding to that used in Fig. 4. The frequencies are averages from measurements on four spectrograms with Hg 4047 as the exciting line and one with Hg 4358 as the exciting line. The 139 lines observed in the  $2700\text{--}3300 \text{ cm}^{-1}$  region belong in the main to the  $\nu_3$  band and the assignments given in Table II were made without much difficulty. The frequencies in parentheses in Table II were not used in the analysis.

#### The O and S Branches

The O and S branches cover a frequency range more than twice as great as that covered by the P and R lines, and are therefore especially suited for the calculation of the molecular constants. The  $O^0$  and  $S^0$  branches terminate in

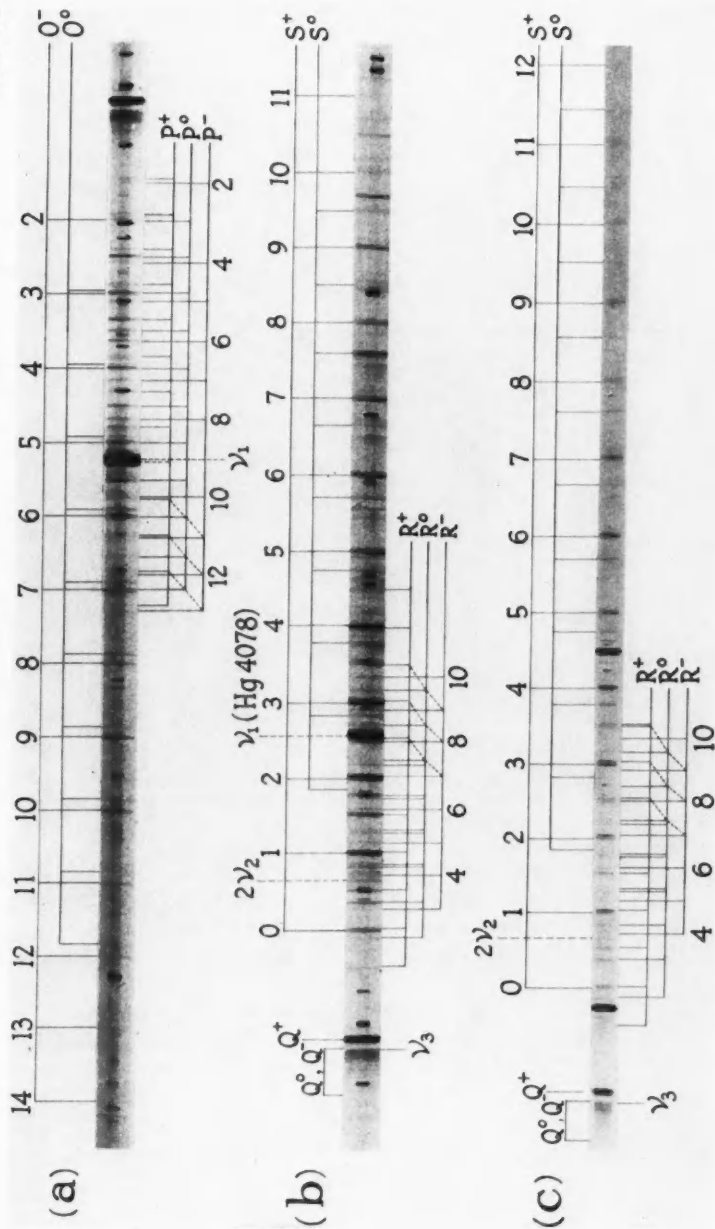
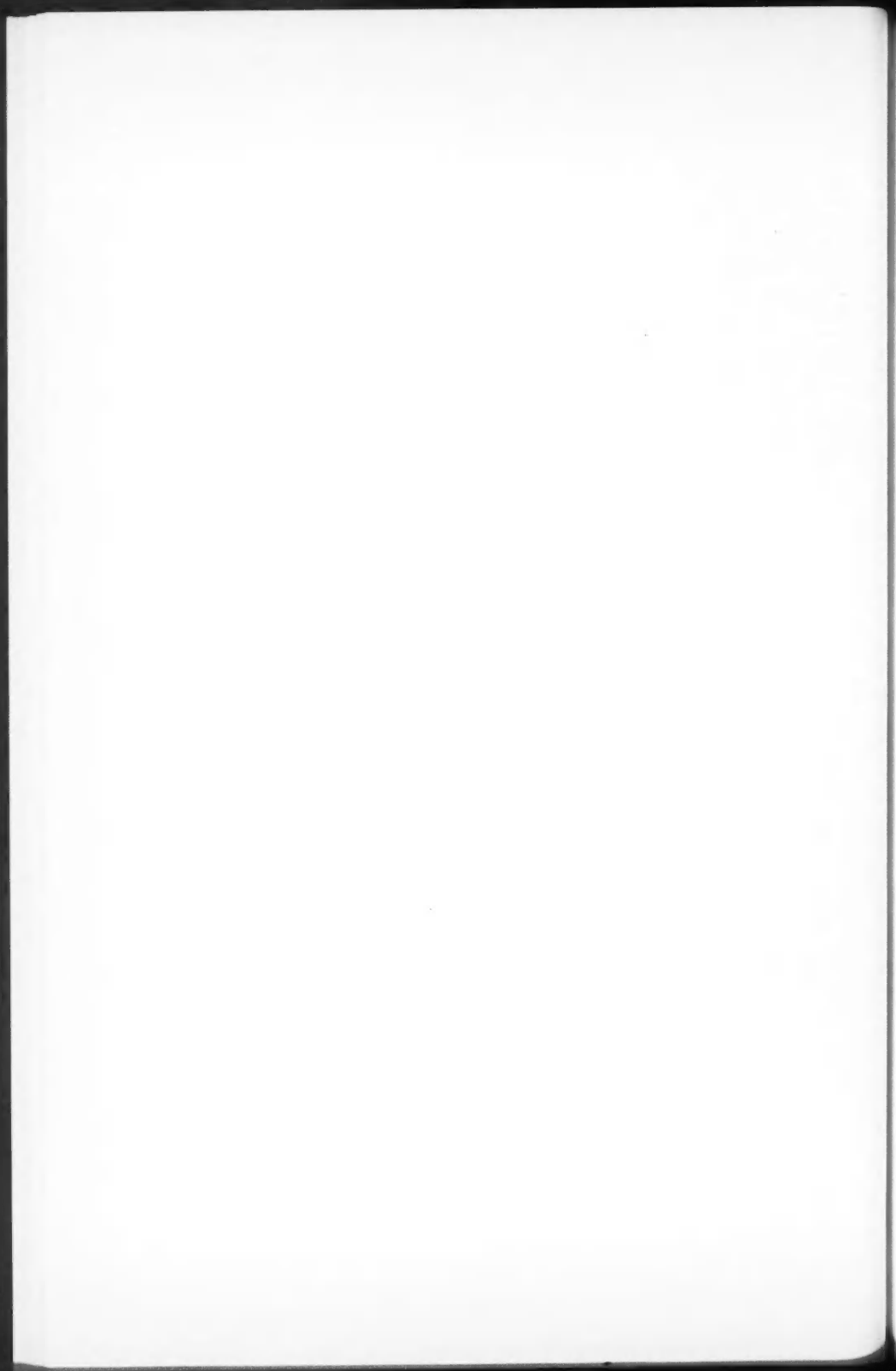


FIG. 3. Spectrograms of the  $\nu_3$  Raman band of methane: (a) the low-frequency side, (b) the high-frequency side of the band excited by Hg 4047, and (c) the high-frequency side excited by Hg 4358.





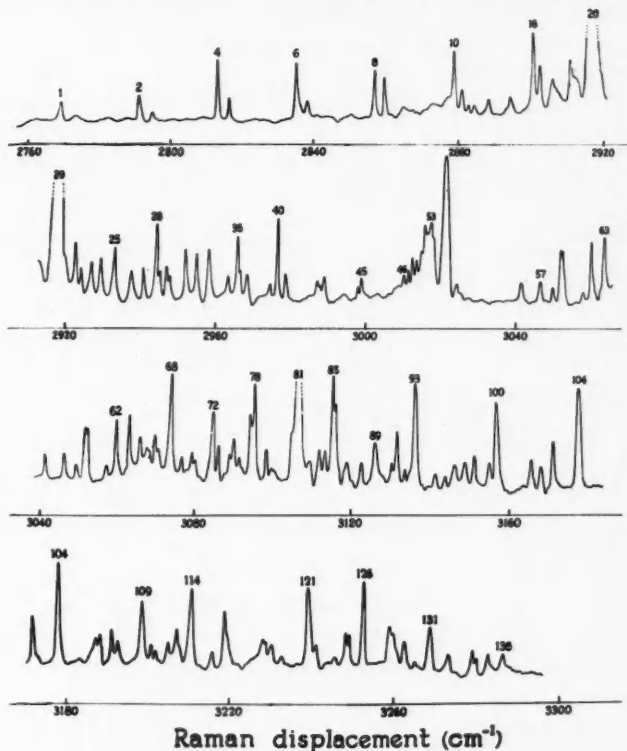


FIG. 4. Microphotometer trace of the 2700-3000  $\text{cm}^{-1}$  region of the Raman spectrum of methane. Most of the maxima belong to the rotational structure of the  $\nu_2$  band.

upper states which are independent of  $\zeta$ ; they were therefore used to obtain the ground state constants. The strong  $O^-$  and  $S^+$  lines seemed to offer the best possibility for the determination of an accurate value of  $\zeta$ . The  $O^+$  and  $S^-$  branches, which should be relatively very weak according to eqs. (6), were not observed, with the possible exception of a weak line at 2946.9  $\text{cm}^{-1}$  which could be assigned to  $O^+(4)$ .

The  $O^0$  and  $S^0$  lines can be represented by eqs. (1) and (2). A least squares calculation with the frequencies of 18 lines, similar to that outlined above for the  $\nu_2$  band, was carried out and gave the following values for the coefficients in eq. (2):

$$\begin{aligned} a &= 3018.95 \pm 0.03, \\ b &= 10.4187 \pm 0.0015, \\ c &= 0.0155 \pm 0.0002, \\ d &= (1.05 \pm 0.04) \times 10^{-4}, \\ e &= (0.024 \pm 0.004) \times 10^{-4}. \end{aligned}$$

TABLE II  
Raman frequencies of the  $\nu_2$  band of  $\text{CH}_4$

Line No.	$\Delta\nu(\text{obs})$ $\text{cm}^{-1}$	Assignment	$\Delta\nu(\text{calc})$ $\text{cm}^{-1}$	Line No.	$\Delta\nu(\text{obs})$ $\text{cm}^{-1}$	Assignment	$\Delta\nu(\text{calc})$ $\text{cm}^{-1}$
—	2725.4	$O^-(14)$	2725.3	53	3017.5	$Q^-(2)$	3017.5
—	2747.2	$O^-(13)$	2747.2	—	(3018.3)	$Q^-(1)$	3018.3
1	2769.1	$O^-(12)$	2769.1	54	3021.4	$Q^+_{\text{max}}$	—
2	2791.1	$O^-(11)$	2791.0	55	3024.2	$O^-(9)(4078)$	—
3	2794.7	$O^0(11)$	2794.8	56	3041.3	$R^+(1)$	3041.4
4	2813.0	$O^-(10)$	2813.0	57	3046.2	$O^-(8)(4078)$	—
5	2816.4	$O^0(10)$	2816.4	58	3049.7	$R^0(2)$	3049.7
6	2834.9	$O^-(9)$	2834.9	59	3051.8	$S^+(0)$	3051.8
7	2838.1	$O^0(9)$	2838.1	60	3052.2	$R^+(2)$	3052.3
8	2856.9	$O^-(8)$	2856.9	61	3057.8	$R^-(3)$	3057.7
9	2859.7	$O^0(8)$	2859.7	62	3059.6	$R^0(3)$	3059.7
10	2878.8	$O^-(7)$	2878.8	63	3063.1	$R^+(3)$	3063.0
11	2881.3	$O^0(7)$	2881.2	64	3065.8	$2\nu_2$	—
12	(2883.3)	$P^-(12)$	2882.5	65	3067.6	$R^-(4)$	3067.1
13	(2884.8)	$P^+(13)$	2884.9	66	3069.5	$R^0(4)$	3069.5
14	2888.5	$P^{0*}(12)$	2888.6	67	(3070.7)	$S^0(1)$	3070.7
15	2895.1	$P^+(12)$	2895.4	—	3073.0	$S^+(1)$	3073.0
16	2900.6	$Q^-(6)$	2900.6	68	3073.7	$R^+(4)$	3073.8
17	2902.7	$O^0(6)$	2902.7	69	3076.6	$R^-(5)$	3076.5
18	(2906.1)	$P^-(10); P^+(11)$	{ 2905.7 2905.9 2910.8	70	3079.2	$R^0(5)$	3079.2
19	2911.2	$P^{0*}(10)$	2910.8	71	3080.0	$R^{0*}(5)$	3080.0
20	2917.0	$\nu_1$	—	—	3084.0	$R^+(5)$	—
21	2922.5	$O^-(5)$	2922.4	72	3084.4	$R^+(5)$	3084.4
22	2924.0	$O^0(5)$	2924.0	73	3085.9	$R^-(6)$	3085.8
23	2926.9	$P^+(9)$	2926.9	74	3088.8	$R^0(6)$	3088.8
24	2929.3	$P^-(8)$	2929.0	75	3089.7	$R^{0*}(6)$	3089.8
25	2933.0	$P^{0*}(8)$	2933.0	76	3091.2	$S^0(2)$	3091.1
26	2937.5	$P^+(8)$	2937.4	77	3094.2	$S^+(2)$	3094.2
27	2940.7	$P^-(7)$	2940.5	78	3095.0	$R^+(6)$	3095.0
28	2944.2	$O^-(4); P^{0*}(7)$	{ 2944.2 2944.0 2945.3	79	3098.3	$R^0(7)$	3098.3
29	2945.3	$O^0(4)$	—	80	3099.5	$R^{0*}(7)$	3099.5
30	2946.9	—	—	—	(3104.4)	$R^-(8)$	3104.1
31	2947.7	$P^+(7)$	2947.8	—	3104.7	$R^{+*}(7)$	—
32	2952.1	$P^-(6)$	2952.0	—	3105.5	$R^+(7)$	3105.5
33	2955.0	$P^{0*}(6)$	2954.9	81	3106.2	$\nu_1(4078)$	—
34	2958.3	$P^+(6)$	2958.2	82	3109.1	$R^{0*}(8)$	3109.2
35	2963.5	$P^-(5)$	2963.4	83	3111.5	$S^0(3)$	3111.4
36	2965.9	$O^-(3); P^{0*}(5)$	{ 2965.9 2965.7 2966.5	84	3113.3	$R^-(9)$	3113.1
37	2966.4	$O^0(3)$	2966.5	85	3115.1	$S^+(3)$	3115.2
38	2968.5	$P^+(5)$	2968.5	86	3115.9	$R^+(8)$	3115.9
39	2974.8	$P^-(4)$	2974.8	87	3118.8	$R^{0*}(9)$	3118.8
40	2976.6	$P^0(4)$	2976.6	88	3122.3	$R^-(10)$	3122.1
41	2978.8	$P^+(4)$	2978.7	—	3124.9	$R^{+*}(9)$	—
42	(2987.2)	$O^-(2); P^{0*}(3)$	{ 2987.5 2987.2 2988.9	—	3125.7	$R^{+*}(9)$	—
43	2988.9	$P^+(3)$	2988.9	89	3126.2	$R^+(9)$	3126.2
44	(2998.2)	$P^{0*}(2)$	2998.0	90	(3130.6)	$O^0(9) \nu_2(4358)$	—
45	2999.0	$P^+(2)$	2999.0	91	3131.7	$S^0(4)$	3131.6
—	3006.3	$Q^-(12)$	3006.1	92	3133.5	$O^-(4)(4078)$	—
—	3007.6	$Q^-(11)$	3007.5	93	3136.2	$S^+(4)$	3136.1
—	3009.0	$Q^-(10)$	3008.8	—	3136.4	$R^+(10)$	3136.4
46	3010.4	$Q^-(9)$	3010.1	94	3141.5	$P^-(6)(4078)$	—
47	3011.6	$Q^-(8)$	3011.4	95	3144.2	$P^0(6)(4078)$	—
48	3012.7	$Q^-(7)$	3012.6	96	3146.5	$R^+(11)$	3146.5
49	3013.8	$Q^-(6)$	3013.7	97	(3149.5)	$O^0(8) \nu_2(4358)$	—
50	3015.0	$Q^-(5)$	3014.8	98	3151.6	$S^0(5)$	3151.6
51	3015.9	$Q^-(4)$	3015.8	99	3155.2	$O^-(3)(4078)$	—
52	3016.6	$Q^-(3)$	3016.7	100	3156.9	$S^+(5)$	3157.0
				101	3165.9	$P^0(4)(4078)$	—
				102	3168.8	$O^0(7) \nu_2(4358)$	—
				103	3171.5	$S^0(6)$	3171.5
				104	3177.8	$S^-(6)$	3177.7

TABLE II (Concluded)

Line No.	$\Delta\nu(\text{obs})$ $\text{cm}^{-1}$	Assignment	$\Delta\nu(\text{calc})$ $\text{cm}^{-1}$	Line No.	$\Delta\nu(\text{obs})$ $\text{cm}^{-1}$	Assignment	$\Delta\nu(\text{calc})$ $\text{cm}^{-1}$
105	3187.3	$\nu_2(4358)$	—	—	3240.2	$S^{++}(9)$	—
106	3188.2	$O^{\beta}(6) \nu_2(4358)$	—	122	3241.9	$P^{\beta}(6) \nu_2(4358)$	—
107	3191.3	$S^{\alpha}(7)$	3191.3	123	3247.1	$R^-(3)(4078)$	—
108	(3192.8)	$P^{\alpha}(11) \nu_2(4358)$	—	124	3248.8	$R^{\alpha}(3)(4078)$	—
109	3198.4	$S^+(7)$	3198.4	125	3249.8	$S^{\alpha}(10)$	3249.8
—	3198.8	$S^{++}(7)$	—	126	3253.1	Grating ghost	—
110	3201.0	$P^{\beta}(11) \nu_2(4358)$	—	127	3259.4	$P^{\alpha}(4) \nu_2(4358)$	—
111	3202.0	$P^{\alpha}(10) \nu_2(4358)$	—	128	3259.7	$S^+(10)$	3259.8
112	3205.1	$Q^-(4) 4078$	—	—	3260.3	$S^{++}(10)$	—
113	3207.3	$Q^-(1)(4078)$	—	129	3262.9	$R^+(4)(4078)$	—
114	3210.8	$S^{\alpha}(8); Q^+_{\text{max}}(4078)$	3210.9	130	3265.7	$R^-(5)(4078)$	—
115	3216.2	$P^{\beta}(9) \nu_2(4358)$	—	131	3269.1	$S^{\alpha}(11)$	3269.0
116	3218.9	$S^+(8)$	3218.9	132	3273.6	$R^+(5)(4078)$	—
—	3219.7	$S^{++}(8)$	—	133	3280.1	$S^+(11)$	3280.1
117	3228.3	$O^{\beta}(4) \nu_2(4358)$	—	134	3280.9	$S^{++}(11)$	—
118	3229.1	$\nu_2(4358)$	—	135	3284.2	$R^+(6)(4078)$	—
119	3230.4	$S^{\alpha}(9)$	3230.4	136	(3288.2)	$S^{\alpha}(12)$	3288.1
120	3232.9	$P^{\beta}(7) \nu_2(4358)$	—	—	3300.4	$S^+(12)$	3300.4
121	3239.3	$S^+(9)$	3239.4	—	3301.3	$S^{++}(12)$	—

From these the following constants were determined:

$$\nu_0 = 3019.00 \pm 0.03 \text{ cm}^{-1},$$

$$B_0 = 5.2406 \pm 0.0011 \text{ cm}^{-1},$$

$$B_1 = 5.1784 \pm 0.0011 \text{ cm}^{-1},$$

$$D_0 = (1.23 \pm 0.07) \times 10^{-4} \text{ cm}^{-1},$$

$$D_1 = (0.86 \pm 0.07) \times 10^{-4} \text{ cm}^{-1}.$$

The values of  $B_0$  and  $D_0$  are very close to those obtained from the  $\nu_2$  band, and the probable error is appreciably less. The value,  $B_0 = 5.2406 \pm 0.0011 \text{ cm}^{-1}$ , for methane is probably the most accurate which has been determined up to the present from either the Raman or the infrared spectrum.

The  $S^+$  and  $O^-$  lines were found by a preliminary analysis to form a regular series. Their frequencies can be represented by a fourth order polynomial in  $m$ , similar to eq. (1), with the coefficients:

$$\begin{aligned}
 a' &= \nu_0 + (3/4)(B_1 - B_0) + 3B_1\zeta_3 - (9/16)(D_1 - D_0), \\
 b' &= B_1 + B_0 + B_1\zeta - (3/2)(D_1 + D_0), \\
 c' &= (1/4)(B_1 - B_0) - (11/8)(D_1 - D_0), \\
 d' &= (1/2)(D_1 + D_0), \\
 e' &= (1/16)(D_1 - D_0).
 \end{aligned}
 \tag{7}$$

A least squares analysis using 24 observed frequencies gave the following values for the coefficients:

$$a' = 3019.72 \pm 0.03,$$

$$b' = 10.7157 \pm 0.0015,$$

$$\begin{aligned}
 c' &= 0.0114 \pm 0.0002, \\
 d' &= (0.93 \pm 0.03) \times 10^{-4}, \\
 e' &= (0.026 \pm 0.002) \times 10^{-4}.
 \end{aligned}$$

The five equations (7) contain six unknown constants; the value of  $B_0 = 5.2406 \text{ cm}^{-1}$ , obtained from the  $O^0$  and  $S^0$  lines, was therefore used to determine the other five:

$$\begin{aligned}
 \nu_0 &= 3018.91 \pm 0.03 \text{ cm}^{-1}, \\
 B_1 &= 5.1950 \pm 0.0011 \text{ cm}^{-1}, \\
 D_0 &= (1.14 \pm 0.05) \times 10^{-4} \text{ cm}^{-1}, \\
 D_1 &= (0.72 \pm 0.05) \times 10^{-4} \text{ cm}^{-1}, \\
 \zeta_3 &= 0.054.
 \end{aligned}$$

The value of  $\nu_0$  obtained here is  $0.09 \text{ cm}^{-1}$  less than that determined from the  $O^0$  and  $S^0$  branches; the difference may be real since it is greater than the sum of the two probable errors. The  $B_1$  differs considerably from that obtained from the  $O^0$  and  $S^0$  branches; this is in qualitative agreement with the conclusions of Stoicheff *et al.* (1952).

A second analysis of the  $O^-$  and  $S^+$  lines was made as follows. The rotational levels,  $F_1^-(J')$  and  $F_1^+(J')$ , of the upper vibrational state were calculated from eq. (3) using the observed frequency shifts and the ground state constants as determined from the  $O^0$  and  $S^0$  branches, but with  $\nu_0 = 3018.91 \text{ cm}^{-1}$ . The following relations should then hold:

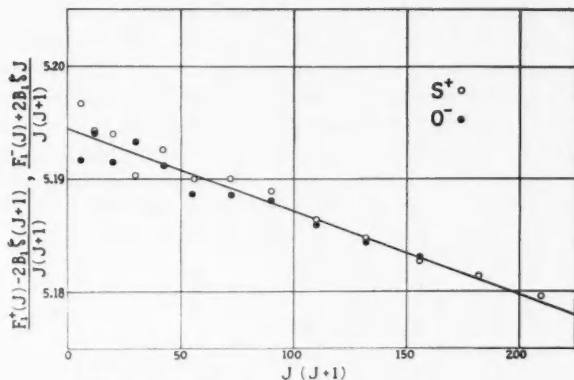
$$\begin{aligned}
 (8) \quad F_1^-(J)/J &= B_1(J+1) - 2B_1\zeta - D_1J(J+1)^2, \\
 F_1^+(J)/(J+1) &= B_1J + 2B_1\zeta - D_1J^2(J+1).
 \end{aligned}$$

A plot of the right-hand sides of these equations against  $J+1$  for the  $O^-$  lines and  $J$  for the  $S^+$  lines gave two parallel lines, slightly curved because of the small cubic term. From the average separation of the two curves the value of  $B_1\zeta = 0.280 \text{ cm}^{-1}$  was found. Equations (8) can be rearranged in the form

$$\begin{aligned}
 (9) \quad \frac{F_1^-(J) + 2B_1\zeta J}{J(J+1)} &= B_1 - D_1J(J+1) \\
 \frac{F_1^+(J) - 2B_1\zeta(J+1)}{J(J+1)} &= B_1 - D_1J(J+1)
 \end{aligned}$$

for the  $O^-$  and  $S^+$  lines, respectively. When the left-hand sides of these equations were plotted against  $J(J+1)$  the graph shown in Fig. 5 was obtained. The least squares regression line gives  $B_1 = 5.194$  and  $D_1 = 0.73 \times 10^{-4} \text{ cm}^{-1}$ ; these are in good agreement with the values obtained by the first method of analysis.

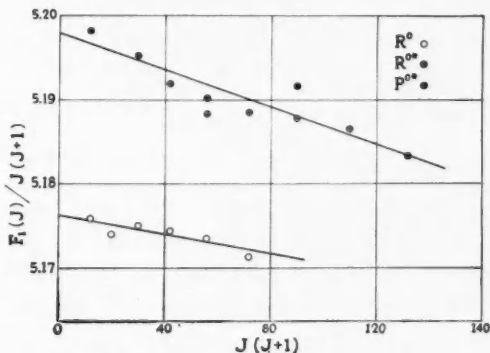
It will be noted that the  $S^+$  lines from  $J = 7$  to  $J = 12$  are resolved into two components. The above analyses were carried out with the stronger components of the doublets. The weak components, with frequency shifts  $0.6$  to  $0.9 \text{ cm}^{-1}$  higher, are designated  $S^{+*}$  in Table II.

FIG. 5. Analysis of the  $O^-$  and  $S^+$  branches of the  $\nu_2$  band.

### The P and R Branches

Some members of each of the  $P^+$ ,  $P^0$ ,  $P^-$ ,  $R^+$ ,  $R^0$ , and  $R^-$  branches were observed. All these series were analyzed by using the ground state constants as determined from the  $O^0$  and  $S^0$  lines, determining the upper state rotational terms from eq. (3) and evaluating the upper state rotational constants from eq. (4) or eqs. (9).

The assignments for the  $P^0$  and  $R^0$  lines given in Table II indicate that the  $P^0$  lines with even values of  $J$  are blended with  $O^-$  lines and that some of the  $R^0$  lines ( $J = 6, 7, 8$ ) are split into two components. The plot of  $F_1^0/J(J+1)$  against  $J(J+1)$  for all of the resolved  $P^0$  and  $R^0$  lines, given in Fig. 6, shows that the points fall on two straight lines. For the  $R^0$  lines with  $J = 3$  to  $J = 8$ , including the *low*-frequency components of the doublets, a least squares analysis gave  $B_1 = 5.176$  and  $D_1 = 0.6 \times 10^{-4} \text{ cm}^{-1}$ . The upper rotational states for these lines are therefore probably the same as for the  $O^0$  and  $S^0$

FIG. 6. Analysis of the  $P^0$  and  $R^0$  branches of the  $\nu_2$  band.

branches. On the other hand, all of the unblended  $P^0$  lines as well as the  $R^0$  lines with  $J = 6$  to  $J = 10$ , including the *high*-frequency components of the doublets, have  $B_1 = 5.198$  and  $D_1 = 1.1 \times 10^{-4} \text{ cm}^{-1}$ ; these lines are designated  $P^{0*}$  and  $R^{0*}$  in Table II. Within the limit of the experimental error these constants are the same as those obtained for the  $O^-$  and  $S^+$  branches.

The corresponding plots for the  $P^-$ ,  $P^+$ ,  $R^-$ , and  $R^+$  branches are shown in Fig. 7. The points for the  $P^-$ ,  $P^+$ , and  $R^-$  lines appear to define the same straight line, for which a least squares computation gave  $B_1 = 5.198$  and

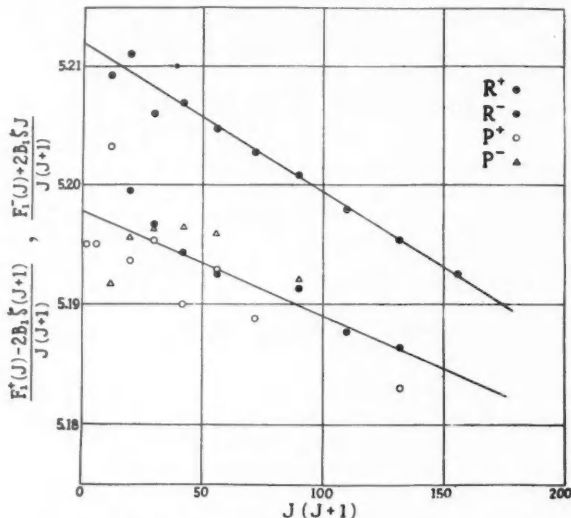


FIG. 7. Analysis of the  $P^+$ ,  $P^-$ ,  $R^+$ , and  $R^-$  branches of the  $\nu_3$  band.

$D_1 = 0.88 \times 10^{-4}$ . These values are again very close to those obtained for the  $O^-$  and  $S^+$  branches. The  $R^+$  lines, however, lie on a distinctly different line for which  $B_1 = 5.212$  and  $D_1 = 1.26 \times 10^{-4} \text{ cm}^{-1}$ . The upper rotational states for these lines are therefore quite different from those for any of the other branches. Table II shows that a few of the  $R^+$  lines have a weak, lower frequency component; these are designated  $R^{+*}$ .

#### The $Q$ Branch

As is evident from Figs. 3 and 4, the most prominent feature of the  $Q$  branch region is the strong maximum at  $3021.4 \text{ cm}^{-1}$ . The region from  $3000$  to  $3020 \text{ cm}^{-1}$  has a good deal of resolved structure which also belongs to the  $Q$  branch. Equations (5) show that the frequency formulae for the  $Q$  branch are:

$$\begin{aligned}
 \Delta\nu_Q^+ &= \nu_0 + (B_1 - B_0)J(J+1) + 2B_1J(J+1) - (D_1 - D_0)J^2(J+1)^2, \\
 \Delta\nu_Q^0 &= \nu_0 + (B_1 - B_0)J(J+1) - (D_1 - D_0)J^2(J+1)^2, \\
 \Delta\nu_Q^- &= \nu_0 + (B_1 - B_0)J(J+1) - 2B_1J(J+1) - (D_1 - D_0)J^2(J+1)^2.
 \end{aligned}
 \tag{10}$$

The maximum at  $3021.4 \text{ cm}^{-1}$  is undoubtedly a band head formed in the  $Q^+$  branch by the negative term  $(B_1 - B_0)J(J+1)$  in the frequency formula overtaking the positive term  $2B_1J(J+1)$ . The position of the head is rather sensitive to the value of  $B_1$ , and it seems probable that the  $Q^+$  branch should be assigned the value  $B_1 = 5.212 \text{ cm}^{-1}$ .

The structure from  $3000$  to  $3020 \text{ cm}^{-1}$  can be assigned very satisfactorily to  $Q^-$  lines with  $B_1 = 5.195 \text{ cm}^{-1}$  and the  $J$  numbering as given in Table II. Assignment to  $Q^0$  lines with  $B_1 = 5.178 \text{ cm}^{-1}$  is less satisfactory. It is more probable that the  $Q^0$  branch should be identified with the region of high intensity on the low-frequency side of the band origin (Fig. 3); a  $B_1$  value of  $5.195 \text{ cm}^{-1}$  would give an intensity distribution which approximates to that observed.

#### Summary of the Rotational Constants

In summarizing the analysis of the  $\nu_3$  band the following points should be emphasized:

(a) The value,  $B_0 = 5.241 \text{ cm}^{-1}$ , as determined from the  $O^0$  and  $S^0$  branches agrees very closely with the value,  $5.240 \text{ cm}^{-1}$ , obtained from the  $\nu_2$  band.

(b) If it is assumed, as seems reasonable, that the remaining branches of the  $\nu_3$  band have the same  $B_0$  value, the analysis shows  $B_1$  values in the range  $5.178$  to  $5.212 \text{ cm}^{-1}$  for the various branches.

(c) Within the accuracy of the frequency measurements the upper state rotational levels are described by three different  $B_1$  values. The distribution of the three values over the branches is summarized in Table III.

TABLE III  
Summary of  $B_1$  values for the  $\nu_3$  band

Branch	$B_1(\text{cm}^{-1})$
$O^0, S^0, R^0$	5.178
$O^-, S^+, P^-, P^{0*}, P^+ \}$	5.195
$R^-, R^{0*}, Q^-, Q^0(?), R^+, Q^+ \}$	5.212

Calculated values of the frequency shifts using the  $B_1$  values of Table III are given in Table II for comparison with the observed shifts.

#### ACKNOWLEDGMENT

The authors are indebted to Professor W. H. Watson, Director of the Computation Center, University of Toronto, for the use of computing facilities.

#### REFERENCES

- BURGESS, J. S., BELL, E. E., and NIELSEN, H. H. 1953. *J. Opt. Soc. Am.* **43**, 1058.  
 FELDMAN, T., ROMANKO, J., and WELSH, H. L. 1955. *Can. J. Phys.* **33**, 138.  
 SHEPHERD, G. G. and WELSH, H. L. 1957. *J. Mol. Spectroscopy*, **1**, 277.  
 STOICHEFF, B. P., CUMMING, C., ST. JOHN, G. E., and WELSH, H. L. 1952. *J. Chem. Phys.* **20**, 498.  
 TELLER, E. 1934. *Hand-und Jahrbuch d. Chem. Physik*, **9**, II.  
 WELSH, H. L., STANSBURY, E. J., ROMANKO, J., and FELDMAN, T. 1955. *J. Opt. Soc. Am.* **45**, 338.

## RELAXATION IN RUBY<sup>1</sup>

R. A. ARMSTRONG AND A. SZABO

### ABSTRACT

The relaxation of the  $(1 \leftrightarrow 2)$  and  $(2 \leftrightarrow 3)$  transitions in chrome-doped  $\text{Al}_2\text{O}_3$  (0.015%) has been studied at  $S$ -band, using a pulsed microwave method, over a range of crystal orientations in the magnetic field at temperatures of  $77^\circ \text{K}$  to  $50^\circ \text{K}$ , and at  $4.2^\circ \text{K}$  and  $1.6^\circ \text{K}$ . A  $T^{-7}$  variation of the relaxation time with temperature was found in the liquid nitrogen range. The relaxation time in this temperature range was found to be independent of crystal orientation, and for the  $(1 \leftrightarrow 2)$  transition was 50 microseconds at  $77^\circ \text{K}$ . At liquid helium temperatures, harmonic cross relaxation was present over most of the range of the crystal orientation studied and was observed at harmonic-to-signal frequency ratios of 2:1, 3:2, and 1:2. The harmonic cross relaxation times were typically 10 to 100 times shorter than the lattice relaxation times, and were independent of temperature. At non-harmonic points at  $4.2^\circ \text{K}$ , the spin-lattice relaxation could be described by one time constant, a value of 300 milliseconds being typical. At harmonic points anomalously long relaxation times as high as 12 seconds were observed.

### INTRODUCTION

A large body of data has been reported in the past twenty-five years on the relaxation process, which tends to maintain paramagnetic ions in thermal equilibrium with their host lattice. Three main experimental methods have been used: the non-resonant (Gorter 1957), the resonant continuous wave (c.w.) (Eschenfelder and Weidner 1953), and the resonant pulse method (Bowers and Mims 1959). In the non-resonant method a relaxation time is measured which is an average over all the spectral lines of the paramagnetic ion. In the resonant method the relaxation of a single line is measured.

The system of spins is usually described by a set of rate equations whereby the time derivatives of the populations of the various levels are given in terms of the spin-lattice transition probabilities. The solution of the rate equations is a sum of exponentials, the number of exponentials being one less than the number of levels, and expressions for the time constants of the exponentials contain all of the transition probabilities in a fairly complicated form.

The pulse method is very convenient for studying the system, since a quantity proportional to the population difference of a pair of levels is measured directly as a function of time, and the return to equilibrium can be resolved graphically into a sum of exponentials. The c.w. method is slower and more indirect since relaxation times are obtained from saturation curves interpreted on the basis of the Bloch equations.

A detailed study of relaxation in 0.015% ruby has been made using the pulse method. The measurements have been made in the range  $T = 50^\circ \text{K}$  to  $77.4^\circ \text{K}$  and at  $1.6^\circ \text{K}$  and  $4.2^\circ \text{K}$ . All of the measurements have been done at 2920 Mc/s. The  $(1 \leftrightarrow 2)$  transition can be observed in the range  $\theta = 0^\circ$  to

<sup>1</sup>Manuscript received May 20, 1960.

Contribution from the Division of Radio and Electrical Engineering, National Research Council, Ottawa, Canada.

Issued as N.R.C. No. 5924.



90° for fields between 500 and 1000 gauss, while observation of the (2↔3) transition is limited by the energy levels and the magnitude of the transition probability to the range  $\theta = 13^\circ$  to  $35^\circ$  in the neighborhood of 3000 gauss (Chang and Siegman 1958).

A single time constant has been observed in the exponential relaxation, except at harmonic points in the liquid helium temperature range. In this range, the coupling between pairs of energy levels whose transition frequencies are harmonically related is large compared with their coupling with the lattice. Consequently, an additional time constant is observed. A set of rate equations has been developed which incorporates the effects of harmonic coupling.

### THEORY

#### *The Rate Equations with Harmonic Cross Relaxation*

The rate equations which describe the dynamics of the energy level populations can be augmented to include the effect of harmonic cross relaxation. As an example of this procedure, consider the case of the three-level spin system shown in Fig. 1 in which  $2(E_1 - E_2) = (E_2 - E_3)$ . In addition to the usual spin-lattice processes, the level populations are also determined by energy-conserv-

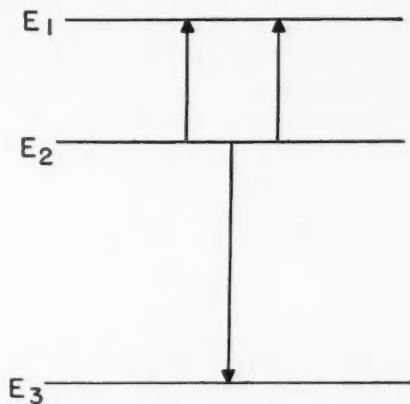


Fig. 1. Energy level diagram showing an energy-conserving triple-spin flip when  $f_{23} = 2f_{12}$ .

ing triple spin flips of the type shown. Consider three ions,  $i, j, k$  (out of a total of  $N$ ), each of which has the energy levels of Fig. 1. The probability per second  $P_{ijk}(2^3; 1^2, 3)$  that two ions undergo the transition (1→2) and one undergoes (3→2) (symbolized by  $(2^3; 1^2, 3)$ ) can be written\*

$$(1) \quad P_{ijk}(2^3; 1^2, 3) = n_1^2 n_3 / N^3 (\omega_{ijk} + \omega_{ikj} + \omega_{kji}) \\ = n_1^2 n_3 / 2N^3 \left( \sum_p P \omega_{ijk} \right)$$

\*Higher order energy-conserving spin flip processes such as  $(2^3, 3^2; 3^2, 1^2)$  and  $(1^3, 3; 2^3, 1)$  will not be considered explicitly. Inclusion of these processes only alters the constant multiplying the cross relaxation matrix (equation (5)) in the linear approximation.

where  $P$  is the permutation operator and  $\omega_{ijk}$  is the probability per second that if ions  $i$  and  $j$  are in level 1, and  $k$  is in level 3, the event  $(2^3; 1^2, 3)$  will occur. In general, for three randomly spaced ions  $(\omega_{ijk} = \omega_{jik}) \neq (\omega_{ikj} = \omega_{kij}) \neq (\omega_{kji} = \omega_{jki})$ . Since there are three combinations (mutually exclusive) of the ions  $i, j, k$  taken two at a time, the probability that one of the three combinations occurs and undergoes the process in question is the sum of the individual probabilities. Note that in equation (1)  $n_m/N$  is the probability that any ion is in state  $m$ . Summing equation (1) over ions, the total number of events  $(2^3; 1^2, 3)$  per second is

$$(2) \quad \begin{aligned} P(2^3; 1^2, 3) &= n_1^2 n_3 / 2N^2 \sum_{i < j < k} \sum_P P \omega_{ijk} \\ &= n_1^2 n_3 / 2N^2 \sum_{jk} \omega_{ijk}. \end{aligned}$$

Assuming that the probability per second for  $(2^3; 1^2, 3)$  equals that of the inverse event  $(1^2, 3; 2^3)$  the rate equations for cross relaxation may be written

$$(3) \quad \left[ \frac{\partial n_1}{\partial t} \right]_{\text{cross}} = 18 \frac{K}{N^2} (n_2^3 - n_1^2 n_3) = -\frac{2}{3} \left[ \frac{\partial n_2}{\partial t} \right]_{\text{cross}} = 2 \left[ \frac{\partial n_3}{\partial t} \right]_{\text{cross}}$$

where

$$K = \sum_{jk} \omega_{ijk} / 18.$$

Linearizing equations (3) by the substitution  $n_j = n_j^0 + \Delta_j$ ,  $(\Delta_j/n_j^0 \ll 1)$  where  $n_j^0$  is the equilibrium population of level  $j$ , we may conveniently write the cross relaxation differential equation in matrix form,

$$(4) \quad \left[ \frac{\partial \Delta}{\partial t} \right]_{\text{cross}} = \mathbf{C} \Delta$$

$$(5) \quad \text{where} \quad \mathbf{C} = K \begin{Bmatrix} -4 & 6 & -2 \\ 6 & -9 & 3 \\ -2 & 3 & -1 \end{Bmatrix}$$

and

$\Delta$  is a column matrix.

The effect of lattice relaxation can also be described by a matrix equation similar to equation (4) so that we have

$$(6) \quad d\Delta/dt = (\mathbf{C} + \mathbf{L}) \Delta = \mathbf{R} \Delta$$

where  $\mathbf{L}$  is a  $3 \times 3$  matrix made up of the usual Boltzmann transition probabilities.

#### *Solution of the Matrix Rate Equation*

A general solution of matrix equations of the type given in equation (6) of dimension  $n$  will first be considered. We observe that if  $\Delta$  is transformed by a matrix  $\mathbf{S}$  to a representation in which  $\mathbf{R}$  is diagonal, then the solution for the transformed  $\Delta$  may be written down readily. A transformation back to the original representation gives the result

$$(7) \quad \Delta = \mathbf{S}^{-1} \mathbf{a} \mathbf{E}$$

where  $\mathbf{a}$  is an  $n \times n$  matrix with  $a_{ij} = \sum_k S_{ik} \Delta_k^i \delta_{ij}$  and  $E$  is a column matrix with  $E_j = \exp(\lambda_j t)$ . Here  $\Delta_j^i = n_j^i - n_j^0$  at zero time, and the  $\lambda_j$ 's are the elements of the diagonal matrix  $\lambda = SRS^{-1}$ . If in the high temperature approximation we take  $\omega_{ij} = \omega_{ji}$  in  $\mathbf{L}$ , then  $\mathbf{R}$  is a symmetric matrix with real eigenvalues and  $\mathbf{S}$  can be chosen to be unitary. A further property of  $\mathbf{S}$  is that  $\sum_j S_{ij} = 0$  for  $\lambda_i \neq 0$ , since  $\sum_j R_{ij} = 0$ .

Since we will want to discuss the behavior of the three-level system of the previous section later, the solution of equation (6) will be obtained for this case. Setting  $\omega_{12} = \omega_{21} = a'$ ,  $\omega_{23} = \omega_{32} = b'$ , and  $\omega_{13} = \omega_{31} = c'$  and defining  $a = a' + 6K$ ,  $b = b' + 3K$ ,  $c = c' - 2K$ ,  $R$  may be written as

$$(9) \quad R = \begin{bmatrix} -(a+c) & a & c \\ a & -(a+b) & b \\ c & b & -(b+c) \end{bmatrix}.$$

Considering only two of the  $\Delta_j$ 's (since they are not all independent) we obtain

$$(10) \quad \begin{Bmatrix} \Delta_1 \\ \Delta_2 \end{Bmatrix} = \frac{1}{\lambda_1 - \lambda_2} \begin{Bmatrix} \Delta_2^i(a-c) - \Delta_1^i(\lambda_2 + a + 2c) & -\Delta_2^i(a-c) + \Delta_1^i(\lambda_1 + a + 2c) \\ -\Delta_1^i(b-a) + \Delta_2^i(\lambda_1 + a + 2c) & \Delta_1^i(b-a) - \Delta_2^i(\lambda_2 + a + 2c) \end{Bmatrix} \begin{Bmatrix} (\exp \lambda_1 t) \\ (\exp \lambda_2 t) \end{Bmatrix}$$

where

$$(11) \quad \lambda_{1,2} = -(a+b+c) \pm \sqrt{a^2 + b^2 + c^2 - ab - ac - bc}.$$

#### Cross Relaxation Time

A cross relaxation time  $\tau_c$  may be defined by equation (11) if the lattice relaxation is neglected. We obtain

$$(12) \quad \tau_c = \left\{ \frac{7}{9} \sum_{jk}^{\text{ions}} \omega_{ijk} \right\}^{-1}.$$

Converting the sum over ions in equation (12) to a sum over sites (Kittel and Abrahams 1953)

$$(13) \quad \tau_c = \left\{ \frac{7}{9} f^2 \sum_{jk}^{\text{sites}} \omega_{ijk} \right\}^{-1}$$

where  $f$  is the fraction of available sites occupied by paramagnetic ions. It is interesting to note that the square-law dependence of  $\tau_c$  on concentration given by equation (13) has been observed for ruby (Mims and McGee 1959).

While a specific calculation of the sum  $\sum_{jk} \omega_{ijk}$  will not be attempted here, it should be noted that if the main contribution to the harmonic transition probabilities arises from the dipolar Hamiltonian (Bloembergen *et al.* 1959), then for computational purposes the above sum should be modified by some sort of averaging over the  $\omega_{ijk}$  since a given  $\omega_{ijk}$  will not be independent of the state of the other ions. This is accomplished formally by replacing the sum  $\sum_{jk} \omega_{ijk}$  with the product of a density of states function  $G_c(\omega_c)$  and a sum over

the dipolar induced transition probabilities  $\omega'_{ijk}$ . Since the dipolar Hamiltonian can only flip two spins at a time, second-order time-dependent perturbation theory must be used to evaluate the  $\omega'_{ijk}$ . A calculation similar to that outlined here has been quantitatively carried out recently (Pershan 1960) describing cross relaxation in LiF by a process involving a simultaneous spin flip of three nuclei. For ruby this type of calculation would be considerably more complex because of its mixed state energy levels.

#### APPARATUS AND MEASUREMENTS

The pulse saturation technique (Davis *et al.* 1958; Bowers and Mims 1959) was used in all the relaxation time measurements. Figure 2 shows the microwave circuitry employed.

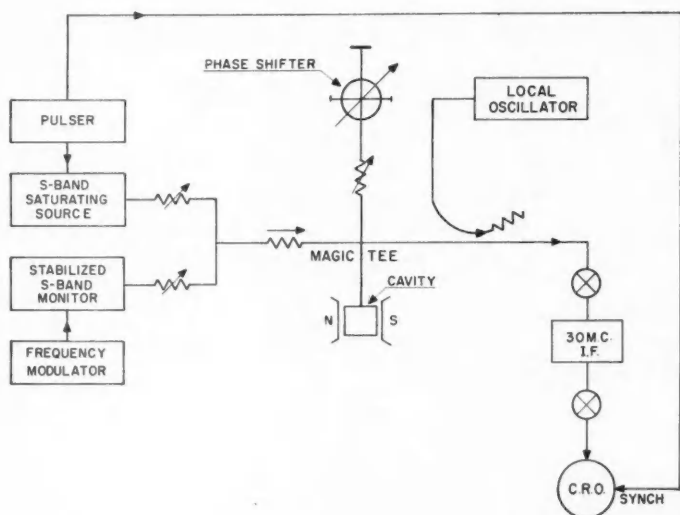


FIG. 2. Microwave circuitry for relaxation time measurements by the pulse saturation technique.

The S-band cavity consisted of a  $\frac{3}{4}$ -wavelength strip mounted centrally in a closed waveguide section of inside dimensions  $\frac{5}{8}$  in.  $\times$   $1\frac{3}{8}$  in.  $\times$   $3\frac{1}{2}$  in. The ruby sample was mounted in polyfoam at the half-wavelength point along the strip. The cavity coupling was adjustable by rotation of the input coaxial line to which the coupling loop was attached. Helium gas sealed in the cavity and waveguide system served as the heat-exchange medium between the ruby and the bath. The cooling liquids used (nitrogen and helium) were excluded from the cavity. Orientation of the ruby sample in the magnetic field could be varied by rotation of the cavity. In the liquid nitrogen range of temperatures

two pieces of ruby  $\frac{1}{2}$  in.  $\times \frac{1}{2}$  in.  $\times \frac{1}{4}$  in. were used, while in the liquid helium range, a single piece  $\frac{1}{2}$  in.  $\times \frac{1}{4}$  in.  $\times \frac{1}{4}$  in. was used. The chromium concentration was 0.015% for all samples.\*

The monitor power entering the cavity arm of the magic- $T$  bridge was always below one microwatt. Spot checks indicated that this level was low enough for the measured relaxation time to be independent of monitor power. If the monitor power is too large, then the observed relaxation time decreases, as is readily seen from a simple consideration of the spin population rate equations (Bloembergen *et al.* 1959). The bridge unbalance (in amplitude) was adjusted so that the relaxation signal fell in the linear range of the receiver. Saturation was produced by a single pulse of  $S$ -band power obtained from a 2K41 klystron delivering approximately 1-watt peak power. The subsequent relaxation signal was displayed on a cathode-ray oscilloscope and photographed.

For the detected signal to be proportional to the magnetic susceptibility, the change of power reflected from the cavity,  $P$ , on and off magnetic resonance, must be small compared with the incident power  $P_i$ . Bowers and Mims (1959) have formulated a linearity condition on  $\Delta S$ , the detected receiver signal corresponding to  $\Delta P$ , for a linear receiver. In the Appendix we consider this problem for a receiver of arbitrary law  $n$  subject to the further condition that the cavity coupling is adjusted for maximum sensitivity. The results show that for a square law or linear receiver,  $\Delta S$  must be less than about  $S_i/10$  in order that the non-linear terms of  $\Delta S$  be less than 10% of the linear term.

#### EXPERIMENTAL RESULTS

Paramagnetic relaxation in ruby (containing 0.015% chromium) has been measured in the liquid nitrogen and liquid helium temperature ranges at 2.92 kMc/s. In the nitrogen range a single time constant is observed. This relaxation time is independent of  $\theta$ , the angle between the crystal axis and the magnetic field, and varies with temperature as  $T^{-7}$ , the temperature dependence predicted for relaxation via Raman processes (Waller 1933). In the helium range the relaxation may be described by the sum of one or two exponentials. The relaxation pattern depends on the angle  $\theta$ . Because of the variation of the time constant with angle it is not possible to give a meaningful temperature dependence in the helium range.

##### *The Liquid Nitrogen Temperature Range*

The relaxation of the populations of the  $(1 \leftrightarrow 2)$  and  $(2 \leftrightarrow 3)$ † transitions were observed at  $15^\circ$  intervals between  $\theta = 0^\circ$  and  $\theta = 45^\circ$ . The relaxation from saturation can be described by a single exponential

$$n_j - n_i = A[1 - \exp(-t/\tau)]$$

No angular dependence was observed. The temperature dependence followed a  $T^{-7}$  law. The relaxation time for the  $(1 \leftrightarrow 2)$  transition was of the order of

\*Analysis by the Department of Mines and Technical Surveys. The figure given is per cent chromium by weight in aluminum oxide.

†The energy levels are numbered as in Chang and Siegman (1958). Level 1 is  $+3/2$  in high field notation and has the highest energy.

two times that of the ( $2 \leftrightarrow 3$ ) transition. This may be a magnetic field effect since the transitions occurred at roughly 800 and 2800 gauss. The results are plotted in Fig. 3.

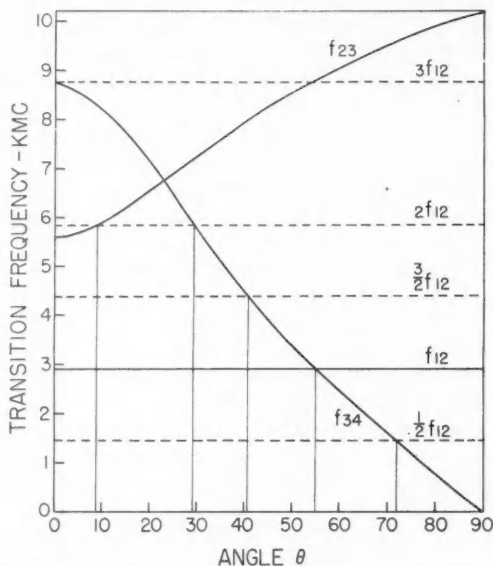


FIG. 3. The temperature dependence of the spin-lattice relaxation time in 0.015% chrome doped  $\text{Al}_2\text{O}_3$  in the liquid nitrogen range.

#### *The Liquid Helium Temperature Range*

The decay from saturation of the ( $1 \leftrightarrow 2$ ) transition was measured at angular intervals of one or two degrees between  $\theta = 0^\circ$  and  $90^\circ$  at temperatures of 4.2° K and 1.6° K. In addition, the ( $2 \leftrightarrow 3$ ) transition was studied over the range  $\theta = 13^\circ$  to  $35^\circ$ ; the limits are set by the variation of the energy levels and the transition probability.

The main feature of the results is the appearance of harmonic cross relaxation (Mims and McGee 1959) at several crystal orientations, as shown in Table I

TABLE I

Harmonic points at 2.92 kMc/s					Amplitude ratio following a 0.2-millisecond pulse		
$\theta$ (obs.)	$\theta$ (theor.)	Signal	Harmonic	$f_h/f_a$	(Theor.)	(Expt.) 4.2° K	(Expt.) 1.6° K
10°	12°	$1 \leftrightarrow 2$	$2 \leftrightarrow 3$	2	8.33	~9	~10
29°	33°	$1 \leftrightarrow 2$	$3 \leftrightarrow 4$	2	4.0	5.1	6.3
41°	40°	$1 \leftrightarrow 2$	$3 \leftrightarrow 4$	$3/2$	2.25	1.5	1.9
72°	74°	$1 \leftrightarrow 2$	$3 \leftrightarrow 4$	$1/2$	0.25	0.4	0.4
16°	19°	$2 \leftrightarrow 3$	$1 \leftrightarrow 2$	2	8.33	5.3	

and Fig. 4. In addition to the 2:1 harmonic points (harmonic frequency: signal frequency) observed by Mims and McGee (1959), we have also observed weaker harmonic couplings at 3:2 and 1:2 points. In the neighborhood of harmonic points, at least two time constants are observed. In certain cases there are probably three time constants present; however, the curves obtained from the oscilloscope photographs have been resolved into two parts only.

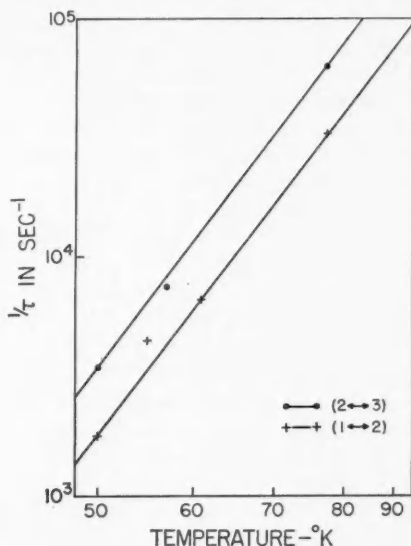


FIG. 4. A plot of the theoretical transition frequencies in ruby with  $f_{12}$  constant at 2920 Mc/s. The observed harmonic points occur at intersections marked by the vertical lines.

The amplitudes of the two exponentials which describe the relaxation at harmonic points after a short (0.2 millisecond) saturating pulse give a measure of the relative heat capacities of the saturated and the harmonic transitions. We define a ratio

$$R = \frac{A_f}{A_s} \approx \frac{T_L}{T_{SH} - T_L}$$

where  $T_L$  is the lattice temperature,  $T_{SH}$  is the spin temperature of the harmonic and saturated transition after the energy sharing is complete,  $A_f$  is the amplitude of the fast decay, and  $A_s$  the amplitude of the slow decay.  $R$  may be calculated readily\* by assuming that the saturated and harmonic transitions share their energy in a time which is short compared with the spin-lattice relaxation time, and that the observed transition is saturated suddenly.

\*As a check on the theory presented earlier,  $R$  was calculated for the case  $(1 \leftrightarrow 2) (2 \leftrightarrow 3)$  harmonic, taking  $K \gg a', b', c'$  in equation (10) giving a result identical with that obtained from the above assumptions.

The former assumption is fully justified in view of the fact that the harmonic cross relaxation time is some 10 to 100 times shorter than the spin-lattice relaxation time. The ratios obtained in this way are compared with experiment in Table I. Similar cross relaxation experiments have recently been reported for nuclear spins in  $\text{NaNO}_3$  by Goldberg (1959).

The long time constant, the short time constant, and the amplitude ratios are plotted as a function of crystal orientation in Fig. 5 for a temperature of  $4.2^\circ \text{K}$ . The results for  $1.6^\circ \text{K}$  are similar, except that the general level of the slow relaxation component is shifted towards longer time constants by an amount which is roughly linear with temperature. The cross relaxation time constants appeared to be independent of temperature.

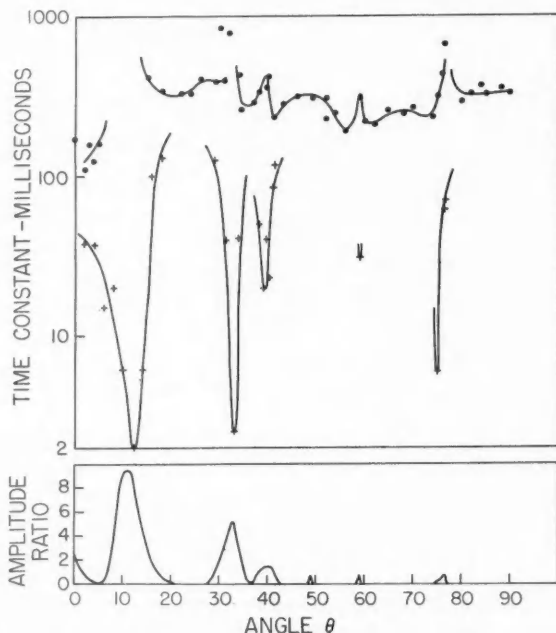


FIG. 5. The dependence of the relaxation time of 0.015% chrome doped  $\text{Al}_2\text{O}_3$  on the crystal orientation at  $4.2^\circ \text{K}$ . In regions where two time constants are observed, the ratio of the amplitudes (fast to slow) of the exponentials describing the decay has been plotted.

An interesting aspect of the results shown in Fig. 5 is that the long relaxation time goes through a maximum at harmonic points. This behavior was investigated in more detail at  $1.6^\circ \text{K}$  for the  $(1 \leftrightarrow 2)$  transition at the  $9^\circ$  harmonic point. It was found that as the length of the saturating pulse was increased from 0.2 millisecond, the amplitude of this long relaxation time increased. (The amplitude of the slow decay for a 0.2 millisecond saturation pulse was too small to obtain a reasonable measure of the time constant; accounting for the blank area at  $9^\circ$  in Fig. 5. The amplitude itself was roughly measurable, however,



to give the entry in Table I.) For a pulse length of 10 seconds or greater, almost the entire decay occurred with the 12-second time constant. The results of long pulse experiments at the  $9^\circ$  and  $33^\circ$  harmonic points are given in Table II.

TABLE II

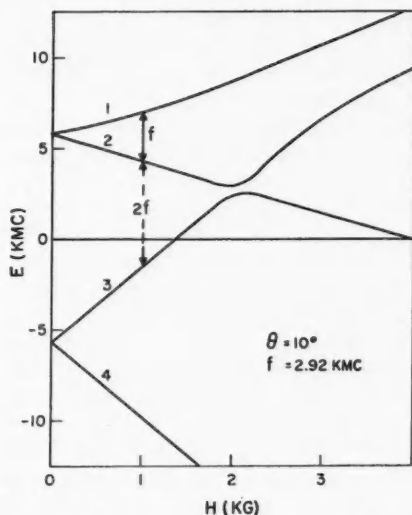
The time constants observed after a 10-second saturating pulse at  $1.6^\circ$  K for transition  $(1 \leftrightarrow 2)$

$\theta$	$7^\circ$	$9^\circ$	$13^\circ$	$31^\circ$	$33^\circ$	$35^\circ$
$\tau$ sec	1.0	12.0	2.1	1.8	1.3	1.9

The slow decay at  $9^\circ$  allowed the following additional experiment to be performed. Just after a long saturation, the crystal orientation was changed rapidly from  $9^\circ$  to  $20^\circ$  and the magnetic field adjusted to bring  $(1 \leftrightarrow 2)$  back on cavity resonance. (The value  $\theta = 20^\circ$  was not an harmonic point.) Immediately after this was done, the susceptibility was observed to be 20% *higher* than the equilibrium value at  $20^\circ$ , and decayed again with a 12-second time constant.

Both the above results can be explained phenomenologically by assuming that the spin-lattice relaxation times,  $\tau_{13}$ ,  $\tau_{23}$ , are long compared with  $\tau_{12}$ . A maser experiment was done which supports, in part, this assumption. At 2830 Mc/s,  $\theta = 13^\circ$  with  $(1 \leftrightarrow 2)$  as the signal and  $(1 \leftrightarrow 3)$  as the pump transition, no maser action could be obtained. A direct measurement of the relaxation indicated that harmonic coupling was not present, hence the relaxation time ratio  $\tau_{12}/\tau_{23}$  was presumably less than 1 (Shulz-duBois *et al.* 1959).

Consider the relaxation at the operating point of  $9^\circ$  shown in Fig. 6. After a long saturation time, the populations of levels 1, 2, 3 are equal. If  $\tau_{13}$ ,  $\tau_{23}$  are long,  $(2 \leftrightarrow 3)$  relaxes to the lattice mainly by transferring energy to  $(1 \leftrightarrow 2)$  by

FIG. 6. Energy levels of  $\text{Cr}^{+++}$  in  $\text{Al}_2\text{O}_3$  at  $\theta = 10^\circ$ .

harmonic coupling. Since the  $(2 \leftrightarrow 3)$  transition has a high heat capacity relative to  $(1 \leftrightarrow 2)$ , the observed relaxation of  $(1 \leftrightarrow 2)$  is slow. This may be shown from equation (11) if we set, in accordance with the above assumptions,  $K \gg a' \gg b', c'$ . We obtain

$$\lambda_{12} \approx -14K \text{ or } -(3/14) a'.$$

Without harmonic coupling the observed relaxation time is  $-2a'$  so that the fractional increase caused by harmonic coupling is given by  $2a'/(3/14)a' = 9\frac{1}{3}$ . If at  $1.6^\circ$  K the relaxation time at non-harmonic points is conservatively taken as 0.5 second, then the above considerations predict a long time constant of about 5 seconds at the harmonic point in question. A calculation of the amplitudes from equation (10) shows that the amplitude of the slow component is the larger, in agreement with the experimental observations.

The rather surprising result of the second experiment can also be qualitatively explained with the same assumptions. After levels 1, 2, 3 have been saturated at  $9^\circ$ , the sum  $n_1 + n_2$  has been increased over the equilibrium value. During the change from  $9^\circ$  to  $20^\circ$ ,  $(1 \leftrightarrow 2)$  has returned to lattice temperature. The observed susceptibility is now larger because of the extra ions trapped in levels 1, 2. These ions relax to 3 via the paths  $1 \leftrightarrow 3$ ,  $2 \leftrightarrow 3$ , which are supposedly slow, hence accounting for the observation of a long time constant again at  $20^\circ$ .

While no physical mechanism is proposed to account for the supposition of the long  $\tau_{23}$ ,  $\tau_{13}$  (in fact, this supposition is contrary to the quadrupole selection rule proposed by Mattuck and Strandberg (1959) giving the spin-phonon transition probabilities for ions of the iron group), it would appear that the recent ENDOR experiments by Terhune *et al.* (1960) may provide the clue to an explanation of the reported time constant behavior. In these experiments long electron-spin time constants are observed which are shown to be associated with the saturation decay of dynamically polarized  $\text{Al}^{27}$  and  $\text{Cr}^{53}$  nuclei. The suggestion is advanced that the host nuclei in ruby have an important role in either the induced transition probabilities or the relaxation mechanism between the electron spin states. A significant difference from our results, however, is that Terhune *et al.* only observe long time constants in the presence of a strong saturating microwave signal.

#### SUMMARY AND CONCLUSIONS

In the temperature range  $T = 50^\circ$  K to  $77.4^\circ$  K the quasi-Raman process apparently governs the relaxation of the spin system to the lattice since the temperature dependence follows the predicted  $T^{-7}$  law.

In the liquid helium temperature range, the temperature dependence is more difficult to measure because the results are more complex. However, the temperature dependence is of the order of  $T^{-1}$  and this indicates that the paramagnetic ions interact with single phonons whose frequencies correspond to the transition frequencies of the ion. Under these conditions the spin-lattice transition probabilities which appear in the rate equations will not, in general, be equal. The solution of the rate equations for a four-level system is then a

sum of three exponentials. However, only a single exponential is observed in the transfer of energy from the spin system to the lattice. (At harmonic points an additional exponential is observed due to energy-sharing within the spin system.) There are several reasons for observing a single exponential: the time constants may be nearly equal and not resolvable experimentally, the transition probabilities may all be equal, or the rate equations themselves may not adequately describe the dynamics of the system.

Harmonic effects profoundly alter the process of relaxation at liquid helium temperatures, the coupling between harmonically related levels being much stronger than their coupling with the lattice even when the ratios of the frequencies are not exactly integral. Harmonic coupling has only recently been considered in maser design (Arams 1960) and explains the failure of a number of otherwise good designs (Bölger *et al.* 1960).

Anomalous long relaxation times have been observed at harmonic points, the longest being 12 seconds at  $\theta = 9^\circ$ . This very long relaxation time can be explained by assuming order of magnitude differences in the spin-lattice transition probabilities.

#### ACKNOWLEDGMENTS

The authors wish to thank Mr. N. W. Broten for his generous advice and help in setting up the microwave circuitry, Mr. E. L. Dimock for technical assistance, and Dr. J. P. Hobson and Mr. P. A. Redhead for many stimulating discussions.

#### APPENDIX

##### MAXIMUM ALLOWABLE SIGNAL FOR LINEAR CAVITY MEASUREMENTS OF SUSCEPTIBILITY

During a measurement of relaxation, the sample susceptibility changes causing a change in cavity dissipation which in turn causes a change in input voltage at the receiver. The change in susceptibility must be limited if the receiver output is to remain proportional to the susceptibility. The maximum signal at the output of an  $n$ -law receiver fulfilling this requirement is derived for the case in which the cavity coupling is adjusted for maximum sensitivity.

Let  $S_d$  be the detected receiver signal corresponding to the power reflected from the cavity when the paramagnetic sample is off magnetic resonance or when the resonance is saturated. Let  $S_0$  and  $S_1$  be similar quantities on magnetic resonance and with the cavity shorted, respectively. In addition, we define

$$A.1(a) \quad \alpha_s = Q_{\text{ext}}/Q_c$$

and

$$A.1(b) \quad \alpha_0 = Q_{\text{ext}} (1/Q_c + 1/Q_{m0})$$

where  $Q_{\text{ext}}$  defines the coupling loss,  $Q_c$  the cavity losses, and  $Q_{m0}$  the unsaturated paramagnetic loss. Immediately after saturation, the paramagnetic loss is zero and builds up to its unsaturated value exponentially with one or more time constants. For simplicity only one time constant will be considered.

We write

$$\text{A.2} \quad \alpha(t) = \alpha_s + (\alpha_0 - \alpha_s) (1 - \exp - \lambda t)$$

since  $Q_m^{-1}(t)$  is proportional to the population difference which goes as  $\exp - \lambda t$ . In terms of the observed signal  $S(t)$

$$\text{A.3} \quad \frac{S(t) - S_0}{S_1} = \left\{ \frac{1 - \alpha(t)}{1 + \alpha(t)} \right\}^n - \left\{ \frac{1 - \alpha_0}{1 + \alpha_0} \right\}^n$$

where  $n$  is the receiver law. Expanding equation A.3 in a binomial series

$$\text{A.4} \quad \frac{S - S_0}{S_1} = \frac{S_0}{S_1} \left\{ \frac{2n(\alpha_0 - \alpha_s)}{1 - \alpha_0^2} \right\} \left\{ \exp - \lambda t + \frac{(n - \alpha_0)(\alpha_0 - \alpha_s)}{1 - \alpha_0^2} \exp - 2\lambda t \dots \right\}.$$

In order that  $S$  reproduce the  $\exp - \lambda t$  variation faithfully, we assume (with Bowers and Mims 1959) the following condition,

$$\text{A.5} \quad \left| \frac{(n - \alpha_0)(\alpha_0 - \alpha_s)}{1 - \alpha_0^2} \right| < 0.1.$$

From equation A.4 to first order

$$\text{A.6} \quad \frac{S_s - S_0}{S_1} = \frac{\Delta S}{S_1} = \frac{S_0}{S_1} \frac{2n(\alpha_0 - \alpha_s)}{1 - \alpha_0^2}$$

so that the condition on  $\Delta S/S_1$  from A.5 and A.6 becomes

$$\text{A.7} \quad \frac{\Delta S}{S_1} < \frac{1}{5} \left| \frac{n S_0}{S_1 (n - \alpha_0)} \right|.$$

The condition for maximum sensitivity may be obtained by a consideration of the expression for the reflected signal

$$\text{A.8} \quad \frac{S}{S_1} = \left\{ \frac{1 - \alpha}{1 + \alpha} \right\}^n$$

where  $\alpha = Q_{\text{ext}}/Q_a$  and  $Q_a$  defines the total loss of the cavity. We want to know the value of  $\alpha$  which gives the maximum change in  $S$  for a given change in  $Q_a$ . This condition is obtained by setting  $(d^2 S)/(d\alpha dQ_a) = 0$  which gives the result

$$\text{A.9} \quad \alpha_{\text{max}} = n \pm \sqrt{(n^2 - 1)}$$

where  $+$ ,  $-$ , correspond to the undercoupled and overcoupled cases, respectively. The condition expressed by equation A.7 on incorporating equation A.9 is

$$\text{A.10} \quad \frac{\Delta S}{S_1} < \frac{n}{5\sqrt{(n^2 - 1)}} \left| \frac{1 - n - \sqrt{(n^2 - 1)}}{1 + n + \sqrt{(n^2 - 1)}} \right|^n.$$

For a linear receiver equation A.10 gives the condition  $\Delta S/S_1 < 0.1$ , while for a square law receiver  $\Delta S/S_1 < 0.08$ .

## REFERENCES

- ARAMS, F. R. 1960. *Proc. I.R.E.* **48**, 108.
- BLOEMBERGEN, N., SHAPIRO, S., PERSHAN, P. S., and ARTMAN, J. O. 1959. *Phys. Rev.* **114**, 445.
- BÖLGER, B., ROBINSON, B. J., and UBBINK, J. 1960. *Physica*, **26**, 1.
- BOWERS, K. D. and MIMS, W. B. 1959. *Phys. Rev.* **115**, 285.
- CHANG, W. S. and SIEGMAN, A. E. 1958. Stanford University Electronics Devices Laboratory, Report TR 156-2.
- DAVIS, C. F., JR., STRANDBERG, M. W. P., and KYHL, R. L. 1958. *Phys. Rev.* **111**, 1268.
- ESCHENFELDER, A. H. and WEIDNER, R. T. 1953. *Phys. Rev.* **92**, 869.
- GOLDBERG, W. I. 1959. *Phys. Rev.* **115**, 48.
- GORTER, C. J. 1957. Progress in low-temperature physics (North Holland, Amsterdam).
- KITTEL, C. and ABRAHAMS, E. 1953. *Phys. Rev.* **90**, 238.
- MATTUCK, R. D. and STRANDBERG, M. W. P. 1959. *Phys. Rev. Letters*, **3**, 369.
- MIMS, W. B. and MCGEE, J. D. 1959. *Proc. I.R.E.* **47**, 2120.
- PERSHAN, P. S. 1960. *Phys. Rev.* **117**, 109.
- SHULZ-DUBOIS, E. O., SCOVIL, H. E. D., and DEGRASSE, R. W. 1959. *Bell System Tech. J.* **38**, 335.
- TERHUNE, R. W., LAMBE, J., MAKHOV, G., and CROSS, L. G. 1960. *Phys. Rev. Letters*, **4**, 434.
- WALLER, I. 1933. *Z. Physik*, **79**, 370.

# ON THE STUDY OF DIFFUSION USING SPIN ECHOES<sup>1</sup>

B. MULLER<sup>2</sup> AND M. BLOOM

## ABSTRACT

The decay of the primary spin echo due to diffusion is governed by the expression  $\exp[-\frac{2}{3}\gamma^2 G^2 D \tau_1^3]$ . This statement is independent of the angles through which the magnetization is rotated by the radio frequency (r-f.) pulses. For the usual range of experimental values, it is also independent of the intensity or homogeneity of the r-f. field, the duration of the r-f. pulse, and the magnitude of the constant field gradient.

## 1. INTRODUCTION

The properties of "spin echoes" in nuclear magnetic resonance have enabled direct measurements of diffusion coefficients to be made in liquids (Hahn 1950; Das and Saha 1954; Carr and Purcell 1954; Torrey 1956; Douglass and McCall 1958). There are, however, still some basic questions which must be answered in order to establish the accuracy of this technique under a variety of experimental conditions. It is the purpose of this paper to answer some of these questions.

In the simplest spin echo experiment two pulses of r-f. separated by a time  $\tau_1$  are applied near the Larmor frequency of the nuclear spins. In addition to nuclear induction signals which are observable immediately following each pulse, a third signal appears having a maximum value at a time  $2\tau_1$  after the first pulse. This signal, called a "spin echo" is due to the rephasing effect of the second pulse on the precession of the spins. As indicated in Fig. 1,  $T_2^*$  is a measure of the time taken for the nuclear spins in the sample to dephase due to the distribution of Larmor frequencies associated with the inhomogeneous external field.

As  $\tau_1$  is varied, the amplitude of the echo  $A(2\tau_1)$  usually decreases monotonically due to loss of "phase memory" by the spins. If the spin system obeys the Bloch (1946) equations, the lengths of the pulses are much less than  $T_2^*$ , the field gradient  $G$  is taken to be uniform, and the motion of the nuclei through  $G$  is described by the diffusion equation, then the amplitude of the echo is given by

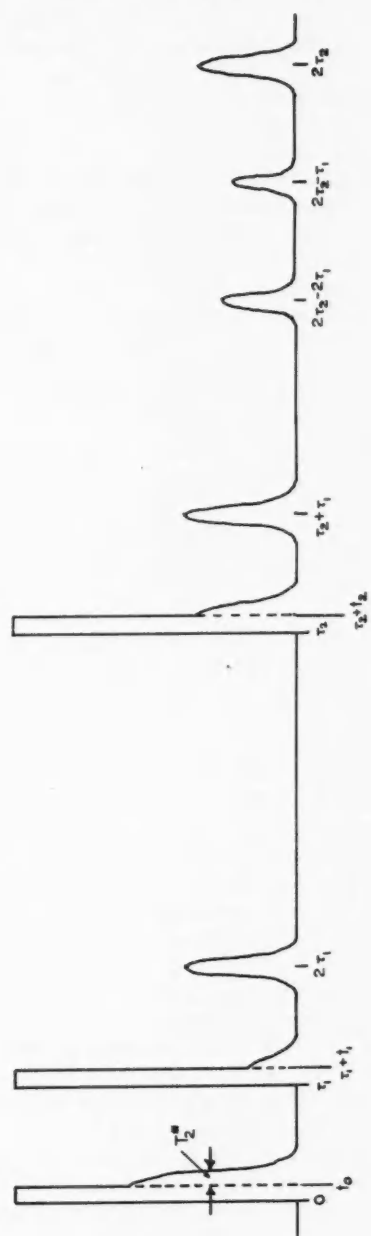
$$(1) \quad A(2\tau_1) \propto \sin \theta_1 \sin^2 \frac{\theta_2}{2} \exp \left[ -\frac{2\tau_1}{T_2} - b\gamma^2 G^2 D \tau_1^3 \right]$$

where  $T_2$  is the "transverse" relaxation time of the system,  $\gamma$  is the gyro-magnetic ratio, while  $\theta_1$  and  $\theta_2$  are the angles through which the first and second pulses respectively rotate the nuclear magnetization vectors.

<sup>1</sup>Manuscript received June 15, 1960.

Contribution from the Department of Physics, University of British Columbia, Vancouver, B.C. Work supported by grants from the National Research Council of Canada.

<sup>2</sup>Holder of U.S. National Science Foundation Science Faculty Fellowship; on leave of absence from University of Wyoming.



The dimensionless factor  $b$  has been calculated by Carr and Purcell (1954) and others (Torrey 1956; Douglass and McCall 1958) to be  $2/3$  for  $\theta_1 = \pi/2$ ,  $\theta_2 = \pi$ . Hahn's original result (Hahn 1950) of  $b = 8/3$  for all  $\theta_1 = \theta_2$  was corrected by Das and Saha (1954), who find  $b = 5/3$ . It is evident that reliable measurements of  $D$  require knowledge of  $b$  for the values of  $\theta_1$  and  $\theta_2$  being used in the experiment and if  $b$  did depend on  $\theta_1$  and  $\theta_2$ , for example, great care would have to be taken in setting up these parameters experimentally.

We have been unable to find any published discussion of the dependence of  $b$  on  $\theta_1$  and  $\theta_2$  or whether the results of Das and Saha on this matter are consistent with results obtained for the special case of a  $\pi/2 - \pi$  pulse train. As is shown explicitly below, these results are not consistent and the difference may be traced to the methods used in treating the statistical properties of the phase accumulated by the spins due to diffusion through the inhomogeneous field. Our results show, in fact, that  $b = 2/3$  for all values of  $\theta_1$  and  $\theta_2$  in agreement with the special case already so extensively studied.

Similar results follow for the "stimulated echo" which appears after the application of a third pulse.

## 2. THEORY

To derive quantitatively the results described above for the effect of spatial diffusion on the spin echo signals the macroscopic magnetization vector of the nuclear spin system must be calculated. An inhomogeneous d-c. magnetic field  $\mathbf{H}(x, y, z)$  is applied in the  $z$ -direction. A field of magnitude  $H_1(t)$ , rotating in the  $x-y$  plane with angular frequency  $\omega$  near the Larmor frequency of the nuclei in the average d-c. field, is applied in the form of pulses, i.e. suppose  $H_1(t) = H_1$  during the pulse and  $H_1(t) = 0$  otherwise. The equation of motion for a magnetization vector  $\mathbf{M}$  in a field  $\mathbf{H}$  is

$$(2) \quad d\mathbf{M}/dt = \gamma\mathbf{M} \times \mathbf{H} + \text{damping terms.}$$

Usually, the damping terms used in calculations of this type are taken to be those given in Bloch's phenomenological equations (Bloch 1946) which assume that in the presence of a steady field in the  $z$ -direction,  $M_z$  approaches its thermal equilibrium value  $M_0$  exponentially with a time constant  $T_1$  while  $M_x$  and  $M_y$  tend towards zero exponentially with a time constant  $T_2$ . These equations are approximately obeyed by many liquids and gases, but for the purpose of calculating effects on  $\mathbf{M}$  due to spatial diffusion in an inhomogeneous magnetic field it will suffice to make two less restrictive assumptions. We will assume that we have waited long enough before the application of the first pulse for the damping terms to have provided us with the thermal equilibrium magnetization  $\mathbf{M}_0$ . We will also make the usual assumption that the decay of the signal caused by diffusion is independent of the relaxation processes due to the damping terms. We can thus ignore relaxation after the first pulse. Relaxation times can, at any rate, always be measured experimentally and if the Bloch equations are valid, our results may be multiplied by damping terms derived elsewhere (Hahn 1950; Das and Saha 1954). Thus our problem is to solve equation (2) for  $\mathbf{M}$  omitting the damping terms and paying special



attention to the effects of the spatial diffusion of spin-containing molecules in the inhomogeneous time-independent magnetic field.

For simplicity, we assume the usual model of a constant field gradient over the entire sample, i.e.  $H(x) = H_0 + Gx$ . This model can usually be well approximated experimentally (Carr and Purcell 1954). Since  $\omega(x) = \gamma H(x)$ , the distribution of Larmor frequencies among the nuclei depends on the geometry of the container (Carr and Purcell 1954). It is described by a normalized distribution function  $g(x)$ :  $\int g(x)dx = 1$ , where  $g(x)dx$  is the fraction of nuclei in the sample between  $x$  and  $x+dx$ .

It is customary, for convenience, to express equation (2) in a co-ordinate system rotating with  $\mathbf{H}_1(t)$ . Then  $\mathbf{H}_1(t)$  is a constant vector along the  $x$ -axis of this co-ordinate system during the pulse. In this co-ordinate system, the Larmor frequency  $\omega(x)$  is transformed to  $\Delta\omega(x) = \omega(x) - \omega = (\gamma H_0 - \omega) + \gamma Gx$ .

We now fix our attention on a given nuclear spin which is located at  $x = x_0$  at  $t = 0$  when the first pulse is applied. Since the nuclear spins are assumed to be noninteracting, there is no loss of generality in replacing the Schroedinger equation for this spin by the equation of motion satisfied by the expectation value of its magnetic moment, which is the same as equation (2) (neglecting damping). If  $U$ ,  $V$ , and  $W$  are the  $x$ ,  $y$ , and  $z$  components of this expectation value in the rotating co-ordinate system, equation (2) becomes

$$\begin{aligned} dU/dt + [\Delta\omega(x(t))]V &= 0 \\ dV/dt - [\Delta\omega(x(t))]U &= -\omega_1 W \\ dW/dt &= \omega_1 V \end{aligned} \quad (3)$$

where

$$\omega_1 = \gamma H_1.$$

We then solve equations (3) for the effect of the r-f. pulses with the spin following a certain path  $x(t)$  in space. To evaluate the  $x$ ,  $y$ , and  $z$  components of magnetization in the rotating co-ordinate system at any time we must first average over all paths up to that time and then sum over all nuclei. The average over all paths is done by assuming that the motion is governed by the diffusion equation, while the sum over all nuclei involves integration over  $g(x_0)$ . Then, if  $U$ ,  $V$ , and  $W$  are expressed in units of the individual magnetic moments of the nuclei in the sample

$$M_x^r = M_0 \int \langle U(x_0, t) \rangle g(x_0) dx_0 \quad (4)$$

where  $M_0$  is the magnetization of the spin system at thermal equilibrium,  $M_x^r$  denotes the  $x$  component of  $\mathbf{M}$  in the rotating co-ordinate system, and the symbol  $\langle \rangle$  denotes an average over all paths of a nucleus initially at  $x = x_0$ . Similar expressions hold for  $M_y^r$  and  $M_z^r$ .

Consider the effects of three pulses of r-f. as indicated in Fig. 1. They are applied at times 0,  $\tau_1$ , and  $\tau_2$  respectively and are of lengths  $t_0$ ,  $t_1$ , and  $t_2$  respectively. The solutions of (3) are simplified if we assume that  $\omega_1$ ,  $1/t_0$ ,  $1/t_1$ , and  $1/t_2$  are all much greater than the spread of Larmor frequencies in the sample. We make this approximation for the moment, but as will be obvious later, our results are not at all dependent on this restriction.

With these assumptions and defining  $\theta_{n+1} = \omega_1 t_n$ ,  $U$ ,  $V$ , and  $W$  at the end of a pulse may be written in terms of their values at the start.

$$\begin{aligned} U(\tau_n + t_n) &= U(\tau_n) \\ (5) \quad V(\tau_n + t_n) &= V(\tau_n) \cos \theta_n - W(\tau_n) \sin \theta_n \\ W(\tau_n + t_n) &= V(\tau_n) \sin \theta_n + W(\tau_n) \cos \theta_n. \end{aligned}$$

In the regions between pulses the solutions are:

$$\begin{aligned} U(t) &= U(\tau_n + t_n) \cos \alpha_{in} - V(\tau_n + t_n) \sin \alpha_{in} \\ (6) \quad V(t) &= V(\tau_n + t_n) \cos \alpha_{in} + U(\tau_n + t_n) \sin \alpha_{in} \\ W(t) &= W(\tau_n + t_n) \end{aligned}$$

where

$$(7) \quad \alpha_{in} = \int_{\tau_n + t_n}^t [\Delta\omega(x(t'))] dt' \cong \int_{\tau_n}^t [\Delta\omega(x(t'))] dt'$$

if  $t_n$  is small enough.

It is convenient, because of the nature of the spin echo experiment, to separate the extra phase  $\beta_{in}$  accumulated by the spins because of their diffusion from the phase which the spins would have accumulated if they had remained fixed at  $x = x_0$  throughout the experiment

$$(8) \quad \alpha_{in} = [\Delta\omega(x_0)] (t - \tau_n) + \beta_{in}.$$

Of course

$$(9) \quad \beta_{in} = \beta_{i0} - \beta_{n0}$$

and

$$(10a) \quad \beta_{i0} = \gamma G \int_0^t [x(t') - x_0] dt'$$

$$(10b) \quad \beta_{n0} = \gamma G \int_0^{\tau_n} [x(t') - x_0] dt'.$$

Since we are here interested only in the echoes, we first study the solutions in the region between the second and third pulses. To simplify notation in what follows we replace the symbol  $\Delta\omega(x_0)$  by  $\Delta\omega$ .

$$(11) \quad \Delta\omega = \gamma H_0 - \omega + \gamma G x_0$$

If at  $t = 0$ ,  $W = 1$ ,  $U = V = 0$ , successive application of (5) and (6) gives, for  $t \geq \tau_1 + t_1$ ,

$$\begin{aligned} U(x_0, t) &= -\sin \theta_1 \sin^2 (\theta_2/2) \sin \{ \Delta\omega(t - 2\tau_1) + (\beta_{i1} - \beta_{10}) \} \\ &\quad + \sin \theta_1 \cos^2 (\theta_2/2) \sin \{ \Delta\omega t + (\beta_{10} + \beta_{i1}) \} \\ &\quad + \cos \theta_1 \sin \theta_2 \sin \{ \Delta\omega(t - \tau_1) + \beta_{i1} \}; \end{aligned}$$

$$\begin{aligned}
 (12) \quad V(x_0, t) = & \sin \theta_1 \sin^2 (\theta_2/2) \cos \{ \Delta\omega(t-2\tau_1) + (\beta_{11}-\beta_{10}) \} \\
 & - \sin \theta_1 \cos^2 (\theta_2/2) \cos \{ \Delta\omega t + (\beta_{10}+\beta_{11}) \} \\
 & - \cos \theta_1 \sin \theta_2 \cos \{ \Delta\omega(t-\tau_1) + \beta_{11} \}; \\
 W(x_0, t) = & -\sin \theta_1 \sin \theta_2 \cos \{ \Delta\omega\tau_1 + \beta_{10} \} + \cos \theta_1 \cos \theta_2.
 \end{aligned}$$

The observable induction signals will be proportional to

$$\omega[(M_x^r)^2 + (M_y^r)^2]^{1/2}.$$

Evidently the integrations over  $x_0$  and the averaging over the extra phase accumulated due to diffusion, which we must perform to evaluate  $M_x^r$  and  $M_y^r$  in equation (4), can be performed independently. Integrals of the type

$$(13) \quad B(t-t_i) = \int \cos [\Delta\omega(t-t_i)] g(x_0) dx_0$$

will be peaked at  $t = t_i$ . For symmetrically shaped samples, the integral of  $\sin [\Delta\omega(t-t_i)]$  is zero. The maximum value of (13) is given by  $B(0) = 1$ . The exact dependence of  $B(t-t_i)$  on  $t$  depends on the shape of the sample, and the rate at which it decays depends on  $G$ . For example (Carr and Purcell 1954), a cylindrical sample of radius  $R$  with the field gradient perpendicular to the cylinder axis gives

$$(14) \quad B(t-t_i) = J_1[\gamma GR(t-t_i)]/\gamma GR(t-t_i)$$

so that  $G$  may be evaluated directly from measurements on the induction signal and the geometry of the sample. Since we are primarily interested in the effect of diffusion on the echo terms, we discuss only terms in (12) which give rise to terms in  $M_x^r$  and  $M_y^r$  peaked at  $t = 2\tau_1$ . For symmetrical samples, the only nonvanishing term of this type is

$$(15) \quad M_y^r(\text{echo}) = M_0 \sin \theta_1 \sin^2 (\theta_2/2) B(t-2\tau_1) \langle \cos (\beta_{11}-\beta_{10}) \rangle.$$

Even if we had an unsymmetrical sample or if we were unable experimentally to place  $\omega$  at the center of a symmetrical distribution and consequently the integrals over odd functions of  $\Delta\omega$  did not vanish, the main features of our result would be unchanged since *all* terms contributing to the echo involve averages over the phase terms of the type

$$(16) \quad \chi_t = \beta_{11} - \beta_{10} = \beta_{10} - 2\beta_{10}.$$

In other words, the rephasing of the nuclear spins at  $t = 2\tau_1$  caused by the second pulse and leading to the echo always causes some cancellation of the phase accumulated due to diffusion in the same way as has been so clearly demonstrated (Carr and Purcell 1954) for the special case  $\theta_1 = \pi/2$ ,  $\theta_2 = \pi$ .

Since  $\chi_t$  is a Gaussian variable (Carr and Purcell 1954; Douglass and McCall 1958), the average over all paths is simply

$$(17) \quad \langle \cos \chi_t \rangle = \int_{-\infty}^{\infty} \cos \chi_t P(\chi_t) d\chi_t = \exp \left[ -\frac{\langle \chi_t^2 \rangle}{2} \right]$$

where  $\langle \chi_i^2 \rangle$  may be calculated from equations (10) assuming  $x(t)$  is governed by the diffusion equation

$$\begin{aligned} (18) \quad \langle \chi_i^2 \rangle &= \langle \beta_{i0}^2 \rangle + 4\langle \beta_{i0}^2 \rangle - 4\langle \beta_{i0} \beta_{i0} \rangle \\ &= (\gamma G)^2 \left[ \frac{2}{3} D t^3 + \frac{8}{3} D \tau_1^3 - 4(D \tau_1^2 t - \frac{1}{3} D \tau_1^3) \right] \\ &= 4(\gamma G)^2 D \left[ \tau_1^3 - \tau_1^2 t + \frac{1}{6} t^3 \right] \end{aligned}$$

as derived by Douglass and McCall (1958). By putting  $t = 2\tau_1$  in (18) and using (15), (10), and  $B(0) = 1$  we see that the amplitude of the echo at  $t = 2\tau_1$  always decays according to

$$(19) \quad M_y^r(2\tau_1) = M_0 \sin \theta_1 \sin^2 (\theta_2/2) \exp \left[ -\frac{2}{3} \gamma^2 G^2 D \tau_1^3 \right].$$

The dependence of the echo amplitude on  $\tau_1$  is thus independent of  $\theta_1$  and  $\theta_2$ .

We have examined this problem of diffusion when two of the approximations made here are not satisfied.

(1) If  $\omega_1$  or  $1/t_n$  are not much larger than the spread of Larmor frequencies in the sample or if  $\omega_1$  is a function of position in the sample, the dependence of the echo on  $\tau_1$  is still identical with (19) as long as  $G$  is constant. This result is to be expected since removal of these restrictions merely means that  $\theta_1$  and  $\theta_2$  depend on  $x_0$ . Thus our integration over  $g(x_0)$  is complicated, but the averaging over the diffusion is not at all affected under usual experimental conditions. A more detailed discussion of this point is given in Appendix II.

(2) If  $G$  is no longer constant, one can solve for the effect of diffusion if the dependence of  $G$  on position is known and one makes the reasonable assumption that each spin-containing molecule experiences a constant value of  $G$  during the experiment. Then the integration over  $x_0$  and the average over all  $x(t)$  are not independent since  $G$  is a function of  $x_0$ . Equation (19) is then replaced by

$$(20) \quad M_y^r(2\tau_1) = M_0 \sin \theta_1 \sin^2 (\theta_2/2) \int \exp \left[ -\frac{2}{3} \gamma^2 G^2 D \tau_1^3 \right] g(x_0) dx_0.$$

#### *The Stimulated Echo*

If a third pulse is applied at  $\tau_2$ , additional echoes appear at  $2\tau_2$ ,  $2\tau_2 - \tau_1$ ,  $2\tau_2 - 2\tau_1$ , and  $\tau_2 + \tau_1$ . The most interesting of these echoes is the latter, which has been called the "stimulated echo". If one continues the type of analysis we have given above, the dependence of the stimulated echo on diffusion can be found in the following way. The only term peaked at  $t = \tau_2 + \tau_1$  involves the product  $\cos \alpha_{i2} \cos \alpha_{i0}$  and we can write

$$(21) \quad M_y^r(\text{stimulated echo}) = \frac{M_0}{2} \sin \theta_1 \sin \theta_2 \sin \theta_3 B[t - (\tau_1 + \tau_2)] \langle \cos \Phi_i \rangle$$

where

$$(22) \quad \Phi_i = \beta_{i2} - \beta_{i0} = \beta_{i0} - \beta_{20} - \beta_{10}$$

and hence following the procedure which gave (18).

$$(23) \quad \langle \Phi_i^2 \rangle = (\gamma G)^2 D \left[ \frac{2}{3} t^3 - 2(\tau_1^2 + \tau_2^2)t + \frac{2}{3} \tau_1^3 + \frac{4}{3} \tau_2^3 + 2\tau_1^2 \tau_2 \right].$$

Therefore the stimulated echo decays according to

$$(24) \quad M_y^r(\tau_1 + \tau_2) = \frac{M_0}{2} \sin \theta_1 \sin \theta_2 \sin \theta_3 \exp \left[ -\frac{\gamma^2 G^2 D}{3} (3\tau_1^2 \tau_2 - \tau_1^3) \right].$$

This result is again independent of  $\theta_1$ ,  $\theta_2$ , and  $\theta_3$  and is in agreement with the result of Herzog and Hahn (1956).

### 3. CONCLUSIONS

The main result we have obtained is that in a spin echo experiment the decays of the primary and stimulated echoes due to spatial diffusion of the spins are independent of the amplitudes and lengths of the r-f. pulses, if the field gradient  $G$  is constant over the entire sample. Thus, in designing a spin echo experiment to measure diffusion coefficients, the greatest care should be taken to produce a uniform  $G$ . It is not at all necessary to satisfy the usual approximations made in the simple theory of spin echoes that  $t_n$ ,  $1/\omega_1 \ll T_2^*$ . These results are obviously useful for a situation in which  $D$  is very small so that very large values of  $G$  must be used to obtain an observable decay of the echo. In such cases it may be impossible to satisfy the above approximations but still simple to produce a relatively constant field gradient over the sample.

### ACKNOWLEDGMENTS

We wish to thank Professor R. E. Burgess and Professor W. Opechowski for some helpful discussions.

### REFERENCES

- BLOCH, F. 1946. Phys. Rev. **70**, 460.  
 CARR, H. Y. and PURCELL, E. M. 1954. Phys. Rev. **94**, 630.  
 DAS, T. P. and SAHA, A. K. 1954. Phys. Rev. **93**, 749.  
 DOUGLASS, D. C. and MCCALL, D. W. 1958. J. Phys. Chem. **62**, 1102.  
 HAHN, E. L. 1950. Phys. Rev. **80**, 580.  
 HERZOG, B. and HAHN, E. L. 1956. Phys. Rev. **103**, 148.  
 TORREY, H. C. 1956. Phys. Rev. **104**, 563.

### APPENDIX I

Since our results are in disagreement with those of Das and Saha (1954) it may be worth while comparing our calculations with theirs. The only difference in our treatments is in the averaging of the phase accumulated by the diffusing spins.

In Das and Saha's paper, their  $\Delta\omega(0)$  corresponds to our  $\Delta\omega(x_0) = \Delta\omega$ . They divide the phase accumulated by the spins between  $t'$  and  $t''$  into three parts instead of two as we do.

$$(25) \quad \int_{t'}^{t''} \Delta\omega(t) dt = \Delta\omega(0)[t'' - t] + \eta_{t'0}[t'' - t'] + \phi_{t't'}$$

where

$$\phi_{t''t'} = \int_{t'}^{t''} \eta_{tt'} dt$$

calling  $\eta_{tt'}$  the change in Larmor frequency by diffusion in the interval  $t-t'$  and  $\phi_{t''t'}$  the extra phase accumulated in the interval  $t''-t'$ . For example, they correctly obtain as the integrand of the expression which yields the primary echo (their equation 19):

$$(26) \quad \sin \xi \sin^2(\xi/2) \cos \Delta\omega(t-2\tau_1) \cos(\phi_{t1}-\phi_{10}) \cos \eta_{10}(t-\tau_1).$$

Now in our notation,  $\eta_{t''t'}(t''-t') + \phi_{t''t'}$  is replaced by  $\beta_{t''t'}$ . With this substitution, it is clear that equation (26) is equivalent to our corresponding equation (15). The recipe for obtaining our equation from theirs is to put  $\eta_{10}$  equal to zero and  $\phi_{t''t'}$  equal to  $\beta_{t''t'}$ . We find that it is not necessary to introduce the  $\eta$ 's. Das and Saha err in averaging independently over  $\phi$  and  $\eta$ . Obviously these are not independent variables since  $\phi$  is a time integral of  $\eta$ .

## APPENDIX II

If we no longer place a limit on the extent of the spread of Larmor frequencies in the sample, the echo terms of the more general solution of equations (3) can be written

$$(27) \quad \begin{aligned} U(x_0, t) &= \frac{\omega_1^2}{\Omega_2^2} \sin^2 \left( \frac{\Omega_2 t_1}{2} \right) \left[ \frac{\omega_1 \Delta\omega}{\Omega_1^2} (1 - \cos \Omega_1 t_0) \cos \{ \Delta\omega(t-2\tau_1) + \chi_t \} \right. \\ &\quad \left. - \frac{\omega_1}{\Omega_1} \sin \Omega_1 t_0 \sin \{ \Delta\omega(t-2\tau_1) + \chi_t \} \right] \\ V(x_0, t) &= \frac{\omega_1^2}{\Omega_2^2} \sin^2 \left( \frac{\Omega_2 t_1}{2} \right) \left[ \frac{\omega_1}{\Omega_1} \sin \Omega_1 t_0 \cos \{ \Delta\omega(t-2\tau_1) + \chi_t \} \right. \\ &\quad \left. + \frac{\omega_1 \Delta\omega}{\Omega_1^2} (1 - \cos \Omega_1 t_0) \sin \{ \Delta\omega(t-2\tau_1) + \chi_t \} \right] \end{aligned}$$

where

$$\begin{aligned} \Omega_1^2 &= \omega_1^2 + (\Delta\omega)^2 \\ \Omega_2^2 &= \omega_1^2 + [\Delta\omega(x(\tau_1))]^2. \end{aligned}$$

The resulting signal will be proportional to

$$(28) \quad \left\langle \frac{\sin^2 \left( \frac{\Omega_2 t_1}{2} \right)}{\Omega_2^2} \cos \chi_t \right\rangle.$$

Since  $\Omega_2$  is not independent of  $\chi_t$ , these cannot in general be averaged separately. However, if

$$(29) \quad \Omega_2^2 - \Omega_1^2 \ll \Omega_1^2$$

we can approximate (28) very well by

$$(30) \quad \frac{\sin^2\left(\frac{\Omega_1 t_1}{2}\right)}{\Omega_1^2} \langle \cos \chi_t \rangle$$

which leads to the same diffusion term as before.

A convenient expression for (29) is obtained if we note that in order to make diffusion measurements with echoes, the rotation angles must be appreciable, i.e.  $\omega_1 t_n \gtrsim 1$  and the values of  $\tau_1^3$  need only be of the order  $3/\gamma^2 G^2 D$  and smaller. Inequality (29) then becomes

$$(31) \quad \gamma^2 G^2 D t_n^3 \ll 1.$$

If we consider typical values of these parameters in e.m.u. for nuclei in a liquid,  $\gamma \simeq 10^4$ ,  $D \simeq 10^{-5}$ ,  $G \simeq 10$ ,  $t_n \simeq 10^{-4}$ ,  $\gamma^2 G^2 D t_n^3 \simeq 10^{-3}$ .

For gases, one might wish to measure a  $D$  as small as  $1 \text{ cm}^2/\text{sec}$ , but even for this case,  $\gamma^2 G^2 D t_n^3 \simeq 10^{-3}$ .

Since inequality (31) is satisfied by a wide range of laboratory conditions, we conclude that the diffusion decay of the echo is governed by the expression  $\exp[-\frac{2}{3}\gamma^2 G^2 D \tau_1^3]$  over the same wide range of experimental conditions.

# BULK PHOTOEFFECTS IN INHOMOGENEOUS SEMICONDUCTORS<sup>1</sup>

C. D. Cox<sup>2</sup>

## ABSTRACT

A photoelectromotive force is observed when a semiconductor is illuminated in the region of a bulk inhomogeneity. A theoretical expression for the bulk photo-e.m.f. is derived which is valid for all levels of illumination in the region of an arbitrary impurity density gradient. With appropriate approximations the complete expression is simplified for the cases of weak and saturation illumination in inhomogeneous extrinsic and nearly intrinsic semiconductors, and in  $p$ - $n$  and  $i$ - $h$  junctions. The theoretical relation between bulk photo-e.m.f. and photoconductive resistance decrease is examined.

Bulk photo-e.m.f. measurements were made on inhomogeneous germanium filaments. These showed the photo-e.m.f. to be linear with weak light illumination. Strong illumination photo-e.m.f. was shown to depend on the impurity distributions adjacent to the illuminated regions. The ratio of bulk photo-e.m.f. to photoconductive resistance decrease was constant at weak illumination in agreement with the theory. Measurements of the photo-e.m.f. as a function of temperature showed a close agreement with the predicted behavior in the near intrinsic range of conductivity.

## INTRODUCTION

In a semiconductor in thermal equilibrium the concentrations of electrons ( $n_0$ ) and of holes ( $p_0$ ) may be expressed in terms of the intrinsic carrier concentration ( $n_i$ ), Fermi level ( $\phi_0$ ), and electrostatic potential ( $\psi_0$ ) (Shockley 1950)

$$n_0 = n_i \exp \frac{q}{kT} (\psi_0 - \phi_0)$$

$$p_0 = n_i \exp \frac{q}{kT} (\phi_0 - \psi_0)$$

where the subscript zero refers to equilibrium conditions. If two regions of a semiconductor in thermal equilibrium have unequal concentrations of ionized impurities, there will be a gradient of carrier density between the regions. Thus from the above equations an impurity density gradient will have associated with it an electrostatic potential gradient or "inner electric field". In the following, "electrostatic potential" refers to conditions in the semiconductor crystal, and "e.m.f." or "voltage" refers to externally observed or applied potential differences.

When light is incident on a semiconductor, a non-equilibrium density of electron-hole pairs is generated. If these excess carriers are generated in the region of an impurity density gradient, the electrons and holes separate under the influence of the inner field resulting in an externally observable e.m.f. known as the bulk photo-e.m.f. Electromotive forces of this type have been reviewed by Tauc (1957).

<sup>1</sup>Manuscript received June 21, 1960.

Contribution from the Physics Department, University of British Columbia, Vancouver, B.C.

<sup>2</sup>Present address: Defence Research Telecommunications Establishment, Defence Research Board, Ottawa, Canada.



The bulk photo-e.m.f. and the illuminated  $p$ - $n$  junction e.m.f. may be considered as two extremes of the same effect. In the case of the bulk photo-e.m.f. the electrostatic potential gradient is gradual, and electrical neutrality may be assumed throughout the crystal. This is not the case in the  $p$ - $n$  junction because of the space charge region associated with the abrupt impurity gradient. Thus although the two effects are fundamentally of the same origin, the analyses of the photo-e.m.f.'s in the two cases are quite different.

### THEORY

Non-equilibrium carrier densities in a semiconductor are expressed by replacing the Fermi level  $\phi_0$  by quasi-Fermi levels or imrefs  $\phi_n$  and  $\phi_p$

$$n = n_0 + \Delta n = n_1 \exp \frac{q}{kT} (\psi - \phi_n)$$

$$p = p_0 + \Delta p = n_1 \exp \frac{q}{kT} (\phi_p - \psi).$$

$\Delta n$  and  $\Delta p$  are respectively the excess concentrations of electrons and holes. It is often convenient to represent a non-equilibrium condition in a semiconductor by an imref and electrostatic potential diagram. This has been done in Fig. 1 for an illuminated bulk inhomogeneity. In the illuminated

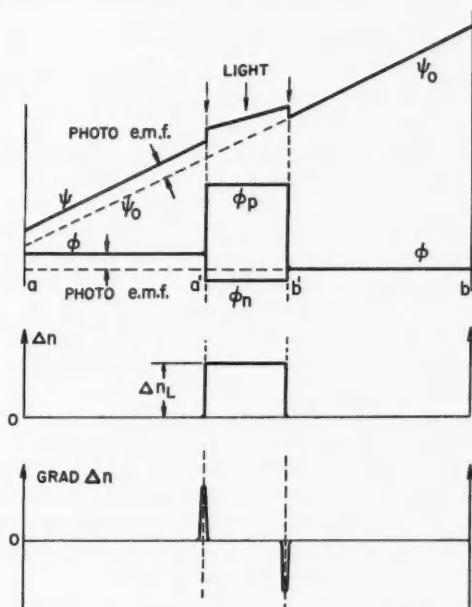


FIG. 1. Assumed carrier distribution in an illuminated bulk inhomogeneity.

region the imrefs separate and take up positions which describe the non-equilibrium carrier densities. At unperturbed points outside the illuminated region, the imrefs return to the equilibrium Fermi level. When incident light gives rise to a difference of Fermi level between points  $a$  and  $b$  on either side of the illuminated region (Fig. 1), the difference is equal to the integral of either imref gradient between the two points.

$$\phi(b) - \phi(a) = \int_a^b \text{grad } \phi_{n,p} dx = \text{e.m.f.}$$

This may be identified with the voltage difference or e.m.f. observed between points  $a$  and  $b$  (Shockley 1950). The electrostatic potential at points  $a$  and  $b$  is at the equilibrium position with respect to the Fermi level. Thus the e.m.f. may also be expressed as

$$\text{e.m.f.} = \int_a^b \text{grad } (\psi - \psi_0) dx.$$

This approach was used by Tauc (1955) in his derivation of an expression for the e.m.f. arising from illumination of an impurity gradient in a semiconductor. The integral is easily evaluated in the case of the illuminated inhomogeneity where electrical neutrality may be assumed. In a  $p$ - $n$  junction, however, the maximum electrostatic potential gradient occurs in the transition region where the crystal is not electrically neutral, and the integral cannot be simply evaluated.

The derivation of an expression for the photo-e.m.f. using the imref integral has the advantage of being more general. The final expression can be applied to an illuminated inhomogeneity or to an illuminated  $p$ - $n$  junction. The analysis of the e.m.f. is made with the following assumptions:

(a) Electrical neutrality is assumed for both illuminated and equilibrium conditions, except in regions of abrupt impurity density gradients such as occur in  $p$ - $n$  junctions. This assumption is only a close approximation to the electrical conditions in the crystal since a tendency towards charge separation is the agency which sets up opposing inner fields, i.e. electrostatic potential gradients. However, except at semiconductor junctions, space charge effects are negligible and the crystal may be assumed electrically neutral (Gunn 1958). The postulate of neutrality implies ambipolar carrier mobility (Rittner 1956), which is taken into account in the analysis and gives rise to factors involving the mobility ratio.

(b) Complete ionization of impurities is assumed. In germanium the activation energies of most impurities are small enough that ionization is essentially complete except at very low temperatures.

(c) Classical statistics are assumed to apply in both the dark and illuminated regions. The most intense illumination used in the experimental studies was not sufficient to approach degeneracy since the added carrier density was estimated to have never exceeded  $10^{16} \text{ cm}^{-3}$ .

(d) The ratio of electron to hole mobility is assumed constant. This assumption holds for all illumination used since hole-electron scattering would not be significant.

(e) Uniform photocarrier generation is assumed across the bulk of the semiconductor. It was not possible to realize this condition in all the experimental work, but from the data it would appear that this assumption is not particularly restrictive. Using this assumption, the theoretical analysis becomes one-dimensional. Thus gradients may be considered vectors lying in the  $x$ -direction, and products or quotients of gradients are scalar quantities with positive values when the gradients are in the same direction, and negative values when they are in opposite directions.

(f) Added carrier concentration is assumed zero outside a region which is ordinarily identified with the illuminated region, and is assumed constant within the illuminated region.

(g) Carrier lifetime is assumed constant, i.e. excess carrier density caused by the illumination is assumed to be a linear function of the light intensity. This assumption is valid for weak illumination, but for large added carrier density the lifetime becomes a function of carrier concentration.

An expression for the bulk photo-e.m.f. is obtained from the integral

$$e_B = \int_a^b \text{grad } \phi_n dx = \int_a^b \text{grad } \phi_p dx$$

by expressing the imref gradients in terms of carrier concentrations and their gradients.

Differentiating the expressions for carrier concentrations

$$\frac{\text{grad } n}{n} = \frac{q}{kT} \text{grad}(\psi - \phi_n)$$

$$\frac{\text{grad } p}{p} = \frac{q}{kT} \text{grad}(\phi_p - \psi).$$

The net current is zero under open circuit conditions, thus

$$bn \text{ grad } \phi_n + p \text{ grad } \phi_p = 0$$

where  $b$  is the ratio of the mobilities of electrons and holes. From these relations we may write

$$-\text{grad } \phi_n = \frac{kT}{q} \frac{1}{n} \frac{\text{grad}(np)}{bn+p}$$

and

$$\text{grad } \phi_p = \frac{bkT}{q} \frac{1}{p} \frac{\text{grad}(np)}{bn+p}.$$

Using the assumption of electrical neutrality,  $\Delta n = \Delta p$  and  $\text{grad } \Delta n = \text{grad } \Delta p$ , the above expressions can be expanded to a form suitable for integration. The integration is carried through using the electron imref.

$$\begin{aligned} e_B &= \int_a^b \text{grad } \phi_n dx = e_a + e_c \\ &= -\frac{kT}{q} \int_a^b \left[ \frac{(n+p) \text{ grad } \Delta n}{n(bn+p)} + \frac{\Delta n \text{ grad}(n_0+p_0)}{n(bn+p)} \right] dx. \end{aligned}$$

The second term of the above expression contains the gradient of equilibrium carrier density. This term is identified with  $e_e$ , the inner field component of e.m.f. arising from the electrostatic potential gradient in the illuminated region. The first term contains the gradient of added carrier density, and is identified with  $e_d$ , the diffusional component of e.m.f.

The origin of the diffusional component may be seen by referring to Fig. 1. As the photo-generated carriers diffuse away from the illuminated region, the electrons, having the greater mobility, tend to move ahead of the holes. This tendency sets up an electric field which slows the electrons and speeds the holes in their movement away from the illuminated region. The field appears as perturbations in the electrostatic potential at the illumination edges. In the case of the illuminated inhomogeneity these perturbations are unequal at the illumination edges, and their difference constitutes the diffusional e.m.f. component.

From assumption (f), the added carrier density is a constant,  $\Delta n_L$ , inside the illuminated region and zero outside it. Thus a significant gradient of added carrier density exists only at  $a'$  and  $b'$ , the edges of illumination, and the first term of the integrand exists only at these points. The equilibrium carrier densities are slowly varying functions of position, and may be considered constant at the two points where the integrand exists. The first term may be written

$$e_d = -\frac{kT}{q} \int_{a'}^{b'} \frac{(n_0 + p_0 + 2\Delta n)}{(n_0 + \Delta n)[p_0 + bn_0 + (b+1)\Delta n]} \frac{d\Delta n}{dx} dx$$

and expanded to

$$e_d = -\frac{kT}{q} \int_{a'}^{b'} \frac{1}{n_0 + \Delta n} \frac{d\Delta n}{dx} dx + \frac{kT}{q} (b-1) \int_{a'}^{b'} \frac{1}{p_0 + bn_0 + (b+1)\Delta n} \frac{d\Delta n}{dx} dx.$$

Using assumption (f), these are evaluated to give

$$e_d = -\frac{kT}{q} \ln \left[ \frac{n_{0a'} + \Delta n_L}{n_{0a'}} \frac{n_{0b'}}{n_{0b'} + \Delta n_L} \right] \\ + \frac{kT}{q} \frac{(b-1)}{(b+1)} \ln \frac{1 + \frac{(b+1)\Delta n_L}{(p_{0a'} + bn_{0a'})}}{1 + \frac{(b+1)\Delta n_L}{(p_{0b'} + bn_{0b'})}}$$

where  $n_{0a'}$ ,  $p_{0a'}$ ,  $n_{0b'}$ ,  $p_{0b'}$  refer to the equilibrium carrier densities at  $a'$  and  $b'$ .

The inner field component of the integrand exists only in the region where the added carrier concentration exists, i.e., between points  $a'$  and  $b'$ . This integral becomes

$$e_e = -\frac{kT}{q} \Delta n_L \int_{a'}^{b'} \frac{\text{grad}(n_0 + p_0)}{n(bn + p)} dx.$$

Using the relation

$$\text{grad } p_0 = -\frac{n_1^2}{n_0^2} \text{grad } n_0$$

$$e_0 = -\frac{kT}{q} \Delta n_L \int_{a'}^{b'} \frac{(n_0^2 - n_1^2)}{n_0(n_0 + \Delta n_L)[bn_0^2 + \Delta n_L(b+1)n_0 + n_1^2]} \frac{dn_0}{dx} dx.$$

This is expanded and evaluated giving

$$e_0 = \frac{kT}{q} \ln \left[ \frac{n_{0b'} n_{0a'} + \Delta n_L}{n_{0a'} n_{0b'} + \Delta n_L} \right] - \frac{2kT}{q} \frac{\Delta n_L(b+1)}{\sqrt{D}} \left[ \tan^{-1} \frac{2bn_0 + \Delta n_L(b+1)}{\sqrt{D}} \right]_{a'}^{b'}$$

where

$$D = 4n_1^2 b - \Delta n_L^2(b+1)^2.$$

The arc tan term takes the tanh form when  $D$  is negative and becomes

$$-\frac{kT}{q} \frac{\Delta n_L(b+1)}{\sqrt{-D}} \left[ \ln \frac{2bn_0 + (b+1)\Delta n_L - \sqrt{-D}}{2bn_0 + (b+1)\Delta n_L + \sqrt{-D}} \right]_{a'}^{b'}$$

When these solutions are applied to the bulk inhomogeneity, the first terms of the diffusional and inner field components cancel. With the remaining terms it is convenient to consider the two extremes of weak and strong illumination, i.e. added carrier density  $\Delta n_L$ , much smaller or much larger than the equilibrium carrier density.

(a) *Weak Illumination* ( $\Delta n_L \ll p_0 + bn_0$ )

The diffusional component reduces to

$$e_d = \frac{kT}{q} (b-1) \Delta n_L \left[ \frac{1}{p_{0a'} + bn_{0a'}} - \frac{1}{p_{0b'} + bn_{0b'}} \right].$$

With the inner field component we assume

$$(b+1)^2 \Delta n_L^2 \ll 4n_1^2 b.$$

Then since  $\tan \theta \doteq \theta$  if  $\theta$  is small

$$\begin{aligned} e_0 &= -\frac{kT}{q} (b+1) \frac{\Delta n_L(n_{0b'} - n_{0a'})}{n_1^2 + bn_{0b'} n_{0a'}} \\ &= \frac{kT}{q} (b+1) \Delta n_L \left[ \frac{1}{p_{0a'} + bn_{0b'}} - \frac{1}{p_{0b'} + bn_{0a'}} \right]. \end{aligned}$$

Thus

$$e_B = e_0 + e_d = \frac{2kT}{bq} \Delta n_L \left[ \frac{1}{n_{0b'}} - \frac{1}{n_{0a'}} \right]$$

in  $n$ -type material where  $bn_0 \gg p_0$ .

Or

$$e_B = e_0 + e_d = -\frac{2bkT}{q} \Delta n_L \left[ \frac{1}{p_{0b'}} - \frac{1}{p_{0a'}} \right]$$

in  $p$ -type material where  $p_0 \gg bn_0$ .

(b) *Strong Illumination* ( $\Delta n_L \gg p_0 + bn_0$ )

The diffusional component becomes approximately

$$e_d = \frac{kT}{q} \frac{b-1}{b+1} \ln \frac{p_{0b'} + bn_{0b'}}{p_{0a'} + bn_{0a'}}.$$

With the inner field component of e.m.f. the log form of the solutions is used, and assuming

$$(b+1)^2 \Delta n_L^2 \gg 4n_i^2 b$$

then

$$e_e = -\frac{kT}{q} \ln \frac{n_{0b'}}{n_{0a'}} = \frac{kT}{q} \ln \frac{p_{0b'}}{p_{0a'}}.$$

Thus

$$e_B = e_e + e_d = -\frac{2}{b+1} \frac{kT}{q} \ln \frac{n_{0b'}}{n_{0a'}}$$

in  $n$ -type material where  $bn_0 \gg p_0$ .

Or

$$e_B = e_e + e_d = \frac{2b}{b+1} \frac{kT}{q} \ln \frac{p_{0b'}}{p_{0a'}}$$

in  $p$ -type material where  $p_0 \gg bn_0$ .

As the bulk photo-e.m.f. increases, the electrostatic potential gradient in the illuminated region decreases to zero. When this condition is reached, no further carrier separation occurs. Thus with strong illumination the bulk photovoltage saturates and becomes independent of the added carrier density.

An expression for bulk photo-e.m.f. in near intrinsic material may be derived by expressing the equilibrium carrier densities in terms of the intrinsic carrier density,  $n_i$ , and the net positive impurity ion density,  $N_D - N_A$ . In near intrinsic material  $n_i \gg N_D - N_A$ , and the equilibrium carrier densities are approximately

$$n_0 = [(N_D - N_A)/2] + n_i$$

$$p_0 = -[(N_D - N_A)/2] + n_i.$$

Substituting these values into the general expression, the photo-e.m.f. becomes

$$e_B = -\frac{2b}{(b+1)^2} \frac{kT}{q} \Delta n_L \frac{(N_D - N_A)_{b'} - (N_D - N_A)_{a'}}{n_i^2}$$

for weak illumination ( $\Delta n_L \ll n_i$ ),

$$e_B = -\frac{2b}{(b+1)^2} \frac{kT}{q} \frac{(N_D - N_A)_{b'} - (N_D - N_A)_{a'}}{n_i}$$

for strong illumination ( $\Delta n_L \gg n_i$ ).

In the above derivations, the analysis is simplified by assumption (f), added carrier density is zero outside the illuminated region and constant within it. This assumption makes it possible to separate the e.m.f. integral into two terms involving different variables. However, when the width of the illuminated region is of the order of a diffusion length, the actual added carrier distribution will be quite different from the assumed rectangular one, and  $n_0$  may change appreciably over the region of significant grad  $\Delta n$ . Under assumption (f) the gradient of added carrier density is a delta function at the illumination edges, whereas in reality a significant gradient exists for a carrier diffusion length from each edge.

The integral identified with the diffusional component of e.m.f. is still valid in this case. However, the equilibrium densities in the integrand are not those at the illumination edges, but are the weighted average of the individual densities over the region of significant grad  $\Delta n$ . In the inner field term, the integral exists over the region of significant added carrier concentration. Although the added carrier concentration is not constant over this region,  $\Delta n_L$  may be considered the average added carrier density over the region of significant  $\Delta n$ . From these considerations, the form of the bulk photo-e.m.f. derived under assumption (f) is valid even though the carrier distribution may be somewhat different from that implied by assumption (f).

Because of carrier diffusion at the edges of illumination, the region of significant  $\Delta n$  may be considerably larger than the illuminated region. Thus the region that is in effect illuminated extends a diffusion length from all sides of the light spot.

#### PHOTO-E.M.F. AT AN ILLUMINATED JUNCTION

The expression for the e.m.f. at an illuminated inhomogeneity cannot be directly applied to an illuminated junction since assumption (a), electrical neutrality, is not valid in the junction transition region. The integral of the imref gradient can be evaluated to a close approximation, however, since the imref change is very nearly zero across the transition region (Fig. 2).

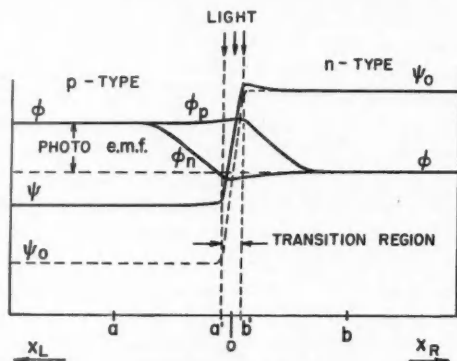
$$\int_a^b \text{grad } \phi_{n,p} dx = \int_a^{a'} + \int_{a'}^{b'} + \int_{b'}^b$$

and

$$\int_{a'}^{b'} \text{grad } \phi_{n,p} dx \doteq 0$$

where  $a$  and  $b$  are two unperturbed points in the  $p$  and  $n$  regions and  $a'$  and  $b'$  mark the edges of the transition layer.

Assuming the impurity density has an abrupt change at a given plane, the width  $W$  of the transition layer may be calculated and is characteristically of the order of  $10^{-4}$  cm wide for impurity density to potential step ratio of  $10^{15} \text{ cm}^{-3} \text{ V}^{-1}$ . The maximum imref slope at an illumination edge is  $kT/qL$  where  $L$  is the diffusion length of the minority carrier. If the imref slope in the transition region approached this value, it would correspond to an imref

FIG. 2. Illuminated *p-n* junction.

change across the layer of no more than  $(kT/q)(W/L)$ . Using typical values of  $W$  and  $L$ , this change is small compared with a typical weak illumination e.m.f. of  $kT/q$  or more. However, the imref slope is zero at one point in the illuminated region, and thus if the light is confined to the transition layer the imref change will be much less than  $(kT/q)(W/L)$ . These considerations show that the assumption of zero imref change in the transition region is justified, and the junction photo-e.m.f. is

$$e_r = \int_a^{a'} \text{grad } \phi_{n,p} dx + \int_{b'}^b \text{grad } \phi_{n,p} dx.$$

Assuming the *p* and *n* regions are homogeneous, the integral of the electron imref gradient gives

$$e_r = -\frac{kT}{q} \ln \left[ \frac{n_{0p} + (\Delta n_L)_p}{n_{0p}} \frac{n_{0n}}{n_{0n} + (\Delta n_L)_n} \right] \\ + \frac{b-1}{b+1} \frac{kT}{q} \ln \frac{1 + \frac{(b+1)(\Delta n_L)_p}{(p_{0p} + bn_{0p})}}{1 + \frac{(b+1)(\Delta n_L)_n}{(p_{0n} + bn_{0n})}}$$

where subscripts *n* and *p* refer to carrier densities in the *n* and *p* sides of the junction. The above expression for photo-e.m.f. was derived with no restrictions as to the type of material on either side of the junction, and can be applied to any type of junction where zero imref change across the transition layer may be assumed.

With weak illumination the added carrier densities are negligible compared with the equilibrium majority carrier densities, and the *p-n* junction photo-e.m.f. becomes

$$e_r = -\frac{kT}{q} \ln \left[ 1 + \frac{(\Delta n_L)_p}{n_{0p}} \right] \\ = -\frac{kT}{q} \ln \left[ 1 + \frac{(\Delta n_L)_n}{p_{0n}} \right].$$



The photo-e.m.f. is proportional to light intensity for very weak illumination,  $\Delta n_L \ll n_{0p}$ ,  $p_{0n}$ , and is proportional to log light intensity with moderate intensities  $n_{0p} < \Delta n_L < p_{0p}$ . With strong illumination the photo-e.m.f. saturates to

$$\begin{aligned} e_r &= -\frac{kT}{q} \ln \frac{n_{0n}}{n_{0p}} + \frac{b-1}{b+1} \frac{kT}{q} \ln \frac{p_{0n} + b n_{0n}}{p_{0p} + b n_{0p}} \\ &= -(\psi_{0n} - \psi_{0p}) + \frac{b-1}{b+1} \frac{kT}{q} \ln \frac{\sigma_{0n}}{\sigma_{0p}}. \end{aligned}$$

The second term of the saturation e.m.f. involves the log of the ratio of the conductivities on either side of the junction, and will ordinarily be small compared with the first term. The saturation e.m.f. is then approximately equal to the equilibrium electrostatic potential step at the junction.

The form of the e.m.f. expression is the same as that derived by Cummerow (1954) for very weak and for moderate light intensities. However, Cummerow's treatment does not include saturation e.m.f.'s.

In a  $p$ - $n$  junction, the second term of the e.m.f. expression will be small since the conductivities of the  $p$  and  $n$  materials will ordinarily be of the same order of magnitude. In this case the only significant contribution to the e.m.f. comes from the first term, which is identical with an expression derived by Fan (1949).

#### RATIO OF BULK PHOTO-E.M.F. TO PHOTOCONDUCTIVE RESISTANCE DECREASE

The bulk photo-e.m.f. observed in an illuminated filament depends on both the light intensity and the impurity density gradient. The added carrier density generated by the light also causes a photoconductive decrease of filament resistance. These two effects originate from the same carrier distribution and their ratio is independent of added carrier density. This gives a convenient method of relating the bulk photo-e.m.f. to the impurity gradient without knowledge of the illumination intensity in terms of the number of effective photons absorbed per second.

The photoconductive resistance decrease is

$$\Delta R = R_0 - \frac{1}{A} \int_L \frac{dx}{\sigma_0 + \Delta \sigma}$$

where  $R_0$  is the filament dark resistance,  $A$  is the cross-sectional area, and the integration is over the entire length,  $L$ , of the filament.  $\Delta \sigma$ , the conductivity increase due to illumination, is assumed small compared to  $\sigma_0$ , the equilibrium conductivity. Thus

$$\Delta R = \frac{q}{A} (\mu_n + \mu_p) \Delta n_L \int_{a'}^{b'} \frac{dx}{\sigma_0^2}.$$

The integration limits are the edges of illumination since  $\Delta n_L$  is zero elsewhere, and  $\sigma_0^2$  becomes  $\sigma_{0L}^2$ , the average conductivity over the illuminated

region. Using the above result and the expression for weak illumination e.m.f. in  $n$ -type material

$$\frac{e_B}{\Delta R} = -\frac{2A}{b+1} \frac{kT}{q} \frac{\sigma_{0a'} - \sigma_{0b'}}{b' - a'}.$$

When the illuminated region is sufficiently narrow

$$\frac{e_B}{\Delta R} = -\frac{2A}{b+1} \frac{kT}{q} \frac{d\sigma_0}{dx}$$

for weak illumination.

#### EXPERIMENTAL

Most of the photo-e.m.f. observations were made with filaments cut from two single crystals of germanium. One of these was a near-intrinsic  $n$ -type crystal of approximately  $40 \Omega \text{ cm}$  resistivity. The other crystal was also  $n$ -type and had a graded resistivity which increased monotonically from  $3 \Omega \text{ cm}$  to  $30 \Omega \text{ cm}$  over a 5-cm length. Typical filaments had a rectangular cross section of  $1 \text{ mm}^2$  and were 1.5 cm long. The ends of the filaments were sandblasted and rhodium-plated. Copper leads were soldered to the plated areas.

A 25-watt zirconium concentrated arc lamp was found to be a convenient light source for photo-e.m.f. work. The arc was demagnified using a two-lens system to provide an intense light spot approximately 0.1 mm diameter. The light intensity at the filament could be decreased by known increments by moving a calibrated neutral density stepped wedge through the light beam. The photo-e.m.f. was measured using a d-c. potentiometer.

The light spot was focussed on a position on the filament which exhibited a large photo-e.m.f. and observations of photovoltage were made as a function of light intensity. Results are shown in Fig. 3 for a near intrinsic filament and for a filament cut from the material having a graded resistivity.

Log photovoltage was plotted against wedge step number, and from the wedge calibration a linear relation between voltage and illumination corresponded to an e.m.f. attenuation of 2.34 db per wedge step, i.e., slope of 1 in log voltage versus log illumination intensity. In all cases the measured slopes were between 0.94 and 1.01. These results verified the linearity of photo-e.m.f. with weak light intensity.

The photo-e.m.f. with strong illumination was different in the two filaments (Fig. 3). In the near intrinsic material, the e.m.f. decreased with increasing strong illumination while in the graded material the e.m.f. increased with light intensity at all levels of illumination. This behavior is explained by diffusion of carriers away from the illuminated region.

The photo-e.m.f. may be considered to have two components. One arises from carriers in the illuminated region and the other from carriers which diffuse away from the edges of the light spot. When the light has reached an intensity sufficient to saturate the photo-e.m.f. component in the illuminated region, a further intensity increase does not affect this part of the e.m.f.

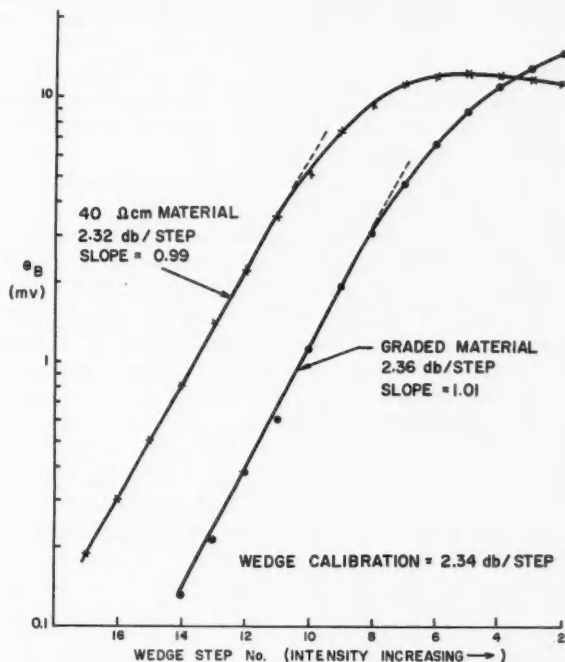


FIG. 3. Dependence of bulk photo-e.m.f. on illumination intensity.

However, an increased number of carriers diffuse from the edges of the light spot into unsaturated regions. The polarity of the e.m.f. component arising from these carriers depends on the impurity distribution. In the graded ingot

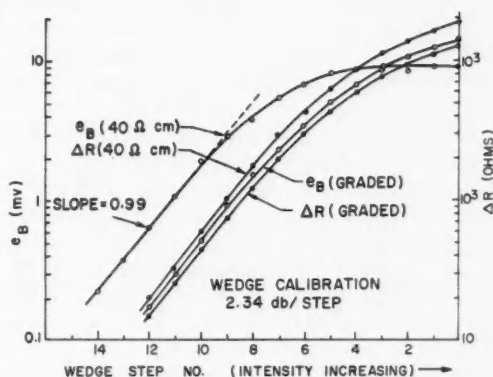


FIG. 4. Dependence of photo-e.m.f. and resistance decrease on illumination intensity.

the impurity density increases monotonically along the filament; thus all e.m.f. components have the same polarity, and the photo-e.m.f. increases at all levels of illumination. In the near intrinsic filament the impurity density fluctuates about a mean value, and in general the diffusional e.m.f. component has a sign opposite to the saturated component, causing the photo-e.m.f. to decrease with increasing strong illumination.

Measurements of photo-e.m.f. and photoconductive resistance decrease are shown in Fig. 4. Electromotive force measurements were made using a d.c. potentiometer, and the resistance decrease was measured by subtracting the resistance of the filament when illuminated from the resistance in the dark. In some cases it was more convenient to measure the ratio  $e_B/\Delta R$  directly using a null method described by Frank (1956). A constant current was passed through the filament and the light beam was chopped by a rotating slotted disk (Fig. 5). The incident light decreased the filament resistance and a voltage pulse was observed at the filament terminals. The amplitude of this pulse was equal to the product of the current and the photoconductive resistance decrease. When the light was incident on a region of impurity density gradient, a photovoltage pulse also appeared at the filament terminals. The direction and magnitude of the current through the filament was adjusted to make the total voltage pulse zero across the filament. At this point, assuming uniform  $\Delta\sigma$  across the filament,

$$e_B/\Delta R = i_t$$

where  $i_t$  was the current through the filament.

The results of Fig. 4 show that the ratio of photo-e.m.f. to resistance decrease is a constant independent of light intensity with weak illumination, as predicted by the theory. With strong illumination the ratio continues essentially constant in the graded filament but not in the 40  $\Omega$  cm filament. This is explained by saturation effects in the illuminated region. With strong illumination the illuminated region approaches zero resistance, and a further filament resistance decrease with increasing strong light intensity is predominantly due to carrier diffusion away from the illuminated region. As a result the resistance decrease which was directly proportional to light intensity with weak light becomes proportional to log light intensity with strong illumination.

The photo-e.m.f. behavior at strong illumination depends on the impurity distribution outside the illuminated region as described above. In the near intrinsic material the photo-e.m.f. decreases with increasing strong illumination, and the ratio  $e_B/\Delta R$  is constant at weak illumination only. With the graded material, the photo-e.m.f. increase at strong illumination depends on diffusional carrier spreading. In this case, the strong illumination dependence of the photo-e.m.f. is similar to that of the resistance decrease, and the ratio of these two quantities is essentially constant at all light intensities.

#### TEMPERATURE DEPENDENCE OF THE BULK PHOTO-E.M.F.

The temperature dependence of the bulk photo-e.m.f. may be calculated

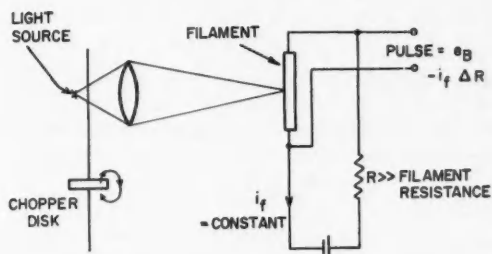
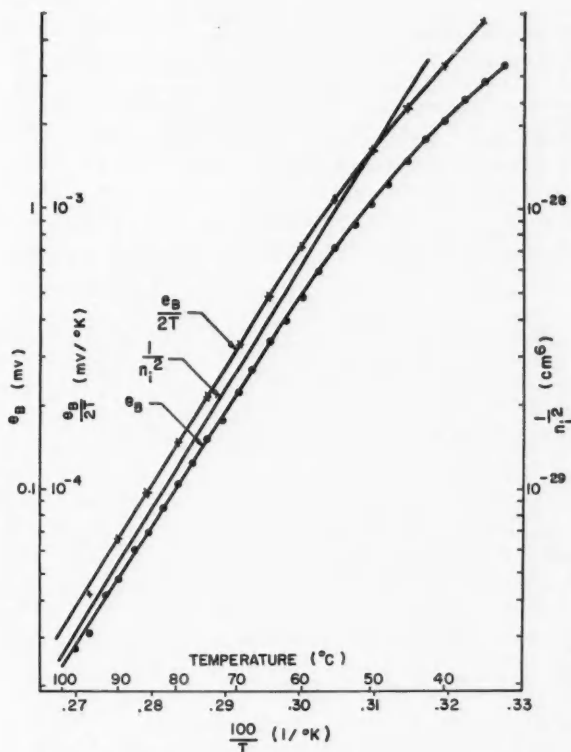
FIG. 5. Measurement of the ratio  $e_B/\Delta R$ .

FIG. 6. Verification of the theory of the photo-e.m.f. temperature dependence in the intrinsic range.

from the e.m.f. expressions derived previously. In the near intrinsic range of conductivity the weak illumination e.m.f. is

$$e_B = -\frac{2b}{(b+1)^2} \frac{kT}{q} \Delta n_L \frac{(N_D - N_A)_v - (N_D - N_A)_a}{n_i^2}.$$

In this expression the temperature dependence of the factor involving mobility may be neglected, and where  $\Delta n_L$  is temperature independent

$$e_B/T \propto 1/n_i^2.$$

Measurements of weak illumination e.m.f. were made with a 40  $\Omega$  cm filament from 20° C to 100° C. Measurements of carrier lifetime over the same temperature range showed a lifetime decrease of approximately 10% as the temperature increased from room temperature to 100° C. Thus the  $\Delta n_L$  factor had negligible effect on the bulk photo-e.m.f. temperature dependence.

In Fig. 6  $\log e_B$ ,  $\log e_B/2T$ , and  $\log 1/n_i^2$  are plotted as functions of reciprocal temperature. The slopes of the  $\log e_B/2T$  and  $\log 1/n_i^2$  plots are equal within experimental error, confirming this feature of the theoretical treatment in the intrinsic range. The value of  $n_i$  was taken from published data (Morin 1954).

In the extrinsic range of conductivity the  $\Delta n_L$  factor becomes the dominant temperature-dependent term. Thus an experimental check of the theoretical temperature dependence of the photo-e.m.f. would require measurements of lifetime as a function of temperature. Such measurements were attempted but the results were not closely reproducible and only qualitative confirmation of the predicted temperature dependence was obtained.

#### ACKNOWLEDGMENTS

I wish to thank Professor R. E. Burgess for his helpful discussions during the course of this work. The assistance of the International Nickel Company of Canada through a Graduate Research Fellowship is gratefully acknowledged. I wish to thank the Defence Research Board for summer assistantships and research facilities provided under grant number 9512-22.

#### REFERENCES

- CUMMEROW, R. L. 1954. Phys. Rev. **95**, 16.
- FAN, H. Y. 1949. Phys. Rev. **75**, 1631.
- FRANK, H. 1956. Czechoslov. J. Phys. **5**, 433.
- GUNN, J. B. 1958. J. Electron. Contr. **4**, 17.
- MORIN, F. J. and MAITA, J. P. 1954. Phys. Rev. **94**, 1525.
- RITTNER, E. S. 1956. Photoconductivity conference (John Wiley & Sons, Inc., New York) p. 215.
- SHOCKLEY, W. 1950. Electrons and holes in semiconductors (D. Van Nostrand Co. Inc., New York).
- TAUC, J. 1955. Czechoslov. J. Phys. **5**, 178.
- 1957. Revs. Modern Phys. **29**, 308.

# THE THERMODYNAMIC DESCRIPTION OF HETEROGENEOUS DISSIPATIVE SYSTEMS BY VARIATIONAL METHODS

## I. A FORMULATION OF THE PRINCIPLE OF MINIMUM RATE OF ENTROPY PRODUCTION WITH APPLICATION TO CERTAIN STATIONARY HETEROGENEOUS CONVECTIVE SYSTEMS<sup>1</sup>

J. S. KIRKALDY

### ABSTRACT

The variational principle

$$\delta \int_{\text{volume}} \sigma dV = 0, \text{ minimum,}$$

where, in terms of the system's independent thermodynamic forces,  $X_i$ , the rate of entropy production per unit volume is

$$\sigma = \sum_{i,k} L_{ik} X_i X_k,$$

is known to describe the integral behavior of certain non-convective, steady (constrained) dissipative processes since the Euler-Lagrange equations corresponding to each degree of freedom define just sufficient steady state conditions to uniquely specify the configuration. It is herein demonstrated that this principle generates the Le Chatelier principle in the presence of constraints which are more general than previously considered. This has made possible the description of certain heterogeneous convective systems which were not previously amenable to unique thermodynamic analysis. In particular, the principle has been used to rationalize cellular or dendritic growth from the melt in alloys. The supporting observation is cited that the non-planar morphology, as compared to a planar one, leads to the maximum rate of conservation of available energy as solute segregation.

### INTRODUCTION

The thermodynamics of steady state dissipative processes has been elucidated in the last few years by Prigogine (1947, 1955), Denbigh (1951), de Groot (1952a), and others. The most important generalization arising from these considerations is that the configuration of certain steady\* dissipative processes can be described as a state of minimum integrated entropy production rate. De Groot (1952b), in an extension of a theorem due to Prigogine, has stated this for a system of interacting discretely uniform parts in which certain of the thermodynamic forces are held constant while the remainder are free to adjust themselves to a steady state defined by the vanishing of their conjugate fluxes. More recently, Brown (1958) has produced a somewhat different construction of the same generality by expressing the entropy production rate for a continuous homogeneous conduction-diffusion-reaction system in terms of the Planck potentials and their gradients. We shall adopt the formulation of de Groot in all our subsequent considerations. In his procedure it was effectively shown that the variational principle of minimum entropy production

<sup>1</sup>Manuscript received March 7, 1960.

Contribution from the Department of Metallurgy and Metallurgical Engineering, McMaster University, Hamilton, Ont.

\*In the following we use the term "steady" to describe a system in which all intensive quantities are temporally invariant at each point of the system. We use the term "stationary" to describe any aged, constrained system. Such a system admits of both steady and oscillatory configurations.

rate generates the correct field equations via the Euler-Lagrange equations, and that the steady configuration is determined by the simultaneous solution of these and the constraint relations. In fact these equations alternatively state the well-known fact that the fluxes conjugate to the free forces must vanish in the steady state.

In the present contribution we examine the approach to and the progress of the stationary state of systems subject to constraints which are rather more general than previously considered with a view to including these systems within the realm of the principle of minimum entropy production. In particular we consider those convective systems for which the boundary conditions or the constraints cannot be stated in the simple fashion proposed by Brown and by de Groot. Furthermore, we propose a formulation which allows us to include certain continuous heterogeneous systems within the variational description. The results are applied to the description of a representative heterogeneous convective metallurgical process.

As an aid to appreciation of the general discussion it is worth while anticipating the ultimate application of the method by examining the experimental situation studied in detail by Tiller and Rutter (1956) and sketched in Fig. 1.

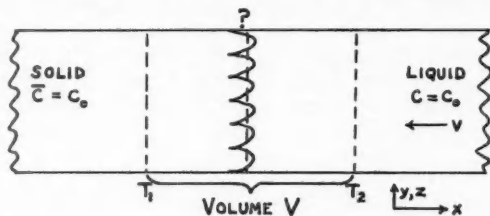


FIG. 1. A schematic advancing steady state solid-liquid interface.

In this setup a liquid alloy of uniform concentration  $C_0$  is fed into a box of volume  $V$  at a constant rate  $v$  and at a temperature  $T_2$  and it emerges from the box at the same rate at a lower temperature  $T_1$  which lies below the freezing range of the alloy. The initial problem is, for given conditions, to define theoretically the morphology of the solid-liquid interface when (and if) the steady state within volume  $V$  has been attained. We shall later consider the possibility of oscillatory (dendritic) configurations.

It should be appreciated at the outset that a continuous theoretical description of the evolution of the steady morphology from some arbitrary initial condition (say a steady configuration with a flat interface) is fundamentally impossible since at least the initial stages of development involve fluctuation phenomena. Even if we could choose an initial condition in which stable morphologic units have been established we would still be faced with the problem of finding a description which explicitly accounts for the correlation between adjacent reaction and diffusion sites along the interface. In the case of homogeneous reactions this correlation is expressed in the fact that the chemical and diffusive processes share the same field variables, the chemical potentials, thus allowing a unique description of transient states by conventional phenomenological means. Unfortunately for the theorist, the sharing of field



variables at an interface is insufficient condition for unique specification of the reaction process. This must be provided by the afore-mentioned information on co-operative phenomena at interfaces.

Since it appears unlikely that suitable advanced techniques for description of this complex transient problem will be shortly forthcoming, we are forced to frame our question in such a way that currently available techniques will suffice. We give up entirely the idea of treating the transient itself and ask if the final steady state can be defined without reference to an actual transient path. As one might expect from knowledge of homogeneous systems, the answer is in the affirmative, and, for the simple system considered here, approximate mathematical procedures can be carried out. This is done by inventing a hypothetical transient path about the steady state for which the chemical and diffusive forces are completely decoupled. It then becomes possible to define the steady state by invoking the Le Chatelier principle and expressing it as a variational principle with variation defined along the afore-mentioned hypothetical path.

In a general discussion of the steady state in terms of thermodynamic forces, the evolution of this state is regarded as occurring within the framework of certain stated constraints and at least one degree of freedom. In the present problem the conditions are such that the concentration entering the box is uniform and fixed at  $C_0$  while the outgoing concentration is only determined to the extent that its mean value at the steady state must equal  $C_0$ , i.e.,  $C(T_2) = C_0$  and  $\bar{C}(T_1) = C_0$ . There exists in this situation a degree of freedom and a single constraint involving interactions between two or more of the independent diffusive forces in the system. However, we cannot "a priori" state the constraint even approximately in an explicit functional form as was done in de Groot's formulation (for example,  $X_i = \text{constant}$ ) but rather we are forced to regard this constraint initially in its most general implicit functional form. It therefore becomes necessary to examine the consequences of the Le Chatelier principle under a more general system of constraints than previously considered.

#### THE EXTENDED PRINCIPLE OF MINIMUM ENTROPY PRODUCTION

Consider a dissipative (thermal conduction-diffusion-reaction) thermodynamic system which can be described by  $n$  forces  $X_i$ , and fluxes  $J_i$ , and let there be  $m$  external constraints which can be stated explicitly as constant single forces or can be expressed explicitly or implicitly by continuous relations involving pairs of forces of the same tensorial and spacial character or scalar components thereof,

$$(1) \quad F_r(X_q, X_{q+1}, x, y, z) = 0, \quad r = 1, 2, 3 \dots m,$$

all the forces and co-ordinates being, in general, implicit functions of the time.\*

\*In a diffusion-reaction system involving interfacial reaction, the diffusive and chemical forces are coupled by their sharing of the chemical potentials as field variables and by certain unknown co-operative phenomena. In our analysis of a particular system, the constraint associated with this coupling will be arbitrarily discarded and the external constraint (1) will be chosen so that the chemical forces can be regarded as independently constrained to constant values.

The constraint equation (1) need not refer to a fixed point in space, but may parametrically follow a line in Cartesian space,  $x(t)$ ,  $y(t)$ ,  $z(t)$ , as implied by the  $t$ -dependence of the co-ordinates. However, it must be constructed in such a way that the magnitude of the volume element associated with each constraint line is invariant. Now, subject to the restriction that interactions occur between pairs of forces or pairs of force components only as in relation (1), we shall demonstrate that the variational principle of minimum rate of entropy production describes a stable steady configuration. The thermodynamic stability of this state is checked by demonstrating that the appropriate Euler-Lagrange equations generate the Le Chatelier principle. For our purposes we state this as follows: In any variation of thermodynamic forces from the steady state along a constraint line the corresponding changes in the thermodynamic fluxes are in such a direction as to moderate the variation.

#### *Description of Steady Configurations*

Let the rate of irreversible entropy production per unit volume be represented as a symmetric, positive definite, quadratic expression in the forces.

$$(2) \quad \sigma = \sum_{i,k}^n L_{ik} X_i X_k$$

(due allowance being made for the tensorial character of the forces in the formation of the products). For the vector fields (diffusion and heat conduction) the  $L$ 's will be assumed constant within any given phase. For the scalar fields (associated with surface chemical reaction) the  $L$ -coefficients will be given by relations of the form

$$(3) \quad L_{ik} = \mathcal{L}_{ik} \delta[x - x'(y, z)]$$

where  $\mathcal{L}_{ik}$  is a constant for a given reaction and the Dirac  $\delta$ -function assumes a non-zero value only at the reacting interface located by the variable  $x'$ . This construction allows the entropy production rate of a surface reaction to be expressed per unit volume. Furthermore, it has no effect on variational procedures carried out on the fluxes and forces since the  $\delta$ -function is a function of the co-ordinates only.

For mathematical completeness we might regard the implicit function  $x'(y, z)$  as containing that information on co-operative phenomena which would make the phenomenological description of the transient process unique. In the absence of this information we plan to choose a convenient transient path via (1) from which the function  $x'(y, z)$  can be calculated. As already noted such a procedure can be used to define the end-point uniquely, but can give us no information about an actual transient path.

The function  $\sigma$  can be represented as a surface in an  $n+1$  dimensional space. A representative two force system is illustrated schematically by the parabolic, positive definite surface of Fig. 2. Consider for illustrative purposes a continuous, homogeneous, non-convective system in which a representative volume element is constrained to configuration A. At time zero we release the constraint on force  $X_1$ , maintaining via fixed boundary values of the field variables a constraint relation,

$$(4) \quad X_2 = \text{constant.}$$

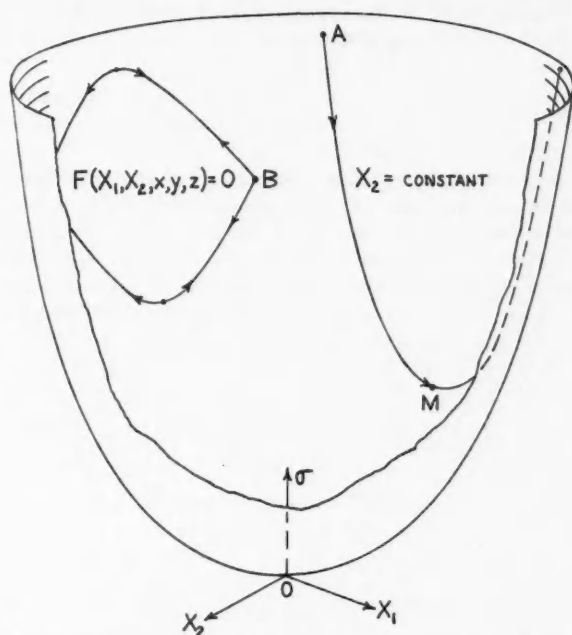


FIG. 2. Schematic two-force  $\sigma$ -surface indicating representative constrained dynamic paths.

Let us tentatively regard  $\sigma$  as a potential function such that the configuration will evolve as far towards the equilibrium configuration

$$(5) \quad \sigma = 0$$

as allowed by the constraint, i.e., to configuration M. This is described mathematically by

$$(6) \quad (\partial\sigma/\partial X_1)_{X_2} = 0$$

or by (2),

$$(7) \quad 2[L_{11}X_1 + L_{12}X_2] = 0.$$

The combination of the simultaneous equations (7) and (4) determines a steady configuration since it infers that both forces are constant. It remains to demonstrate that the configuration is a stable one. De Groot (1952c) has done this by showing that (7) in combination with the flux equations,

$$(8) \quad J_i = \sum_k L_{ik}X_k$$

generates the Le Chatelier principle. Alternatively, one can obtain a direct indication of the stability of the state by noting that (7) specifies the vanishing of the flux conjugate to the free variable, viz.,

$$(9) \quad J_1 = 0.$$

If every volume element in the system follows a similar path then (6) can be regarded as an Euler-Lagrange equation and we can construct the variational relation

$$(10) \quad \delta \int_{\text{volume}} \sigma dV = 0, \text{ minimum.}$$

This simplified treatment contains the essence of Prigogine and de Groot's formulation of the principle of minimum entropy production.\* It can be applied approximately to the thermal diffusion cell (a materially closed system) where in principle we can constrain the thermal gradients to an approximately constant value allowing the free diffusive forces to approach a steady configuration defined by the vanishing of the conjugate diffusive flux.

In a more general multiforce system with given constraints (1) and at least one degree of freedom, the general tendency of the whole system will still be towards the equilibrium configuration and the constraint lines will still pass through extrema in the  $\sigma$ -surface. However, it may not be possible, because of the internal constraints (the mass balance, constitutional requirements, etc.), for all such lines to attain minima. Figure 2 represents this situation for two neighboring volume elements initially constrained to configuration B. If all volume elements attain stable extrema after the same infinite time then Euler-Lagrange equations of type (6) evaluated along the appropriate constraint lines exist for each degree of freedom in the system. Although some of these may define maxima, most will define minima and the integral construction (10) still applies. The minimal nature of the extremum in such mixed cases follows from the positive definite character of the integral entropy production, although the rigorous proof of this lies in the realm of higher mathematics (Kimball 1952).

A typical Euler-Lagrange equation for constraints of type (1) can be written in the following alternative forms:

$$\begin{aligned} (11) \quad \frac{1}{2}d\sigma &= \left[ L_{11}X_1 + L_{12}X_2 + L_{12}X_1 \frac{dX_2}{dX_1} + L_{22}X_2 \frac{dX_2}{dX_1} \right] dX_1 \\ &= J_1 dX_1 + J_2 dX_2 \\ &= X_1 dJ_1 + X_2 dJ_2 = 0, \end{aligned}$$

with variation along the particular constraint line

$$(12) \quad F(X_1, X_2, x, y, z) = 0.$$

The fluxes used in (11) are those defined by (8) and therefore do not include any of the non-dissipative convective flow terms.

\*It should be pointed out here that constraint equation (4) can be regarded as exact for incremental variations about the extremum only, since fixed boundary values imply this to the first order only. In Brown's formulation (1958), where the constraints are stated explicitly as boundary values, the same extremum is obtained as here. In our own formulation of the principle of minimum entropy production which follows, the precise form of the constraints is immaterial since they are stated in a general implicit form (cf. relation (12)).

Differentiating (12) gives

$$(13) \quad dF = (\partial F/\partial X_1)dX_1 + (\partial F/\partial X_2)dX_2 = 0$$

which in combination with (11) gives

$$(14) \quad J_1 = J_2[(\partial F/\partial X_1)/(\partial F/\partial X_2)].$$

The solution of the simultaneous equations (12) and (14) represents a steady configuration in which no flux associated with a degree of freedom in the system need vanish. This is characteristic of convective systems in which the non-dissipative fluxes continually prime the dissipative ones to maintain the steady state. It remains to demonstrate the thermodynamic stability of the configuration defined by (10), (11), or (14).

#### *Stability of the Generalized States of Minimum Entropy Production*

The essential problem is to demonstrate that a steady state of minimum rate of entropy production constrained as in (12) generates the Le Chatelier principle and therefore describes a stable configuration. Let the steady configuration be represented by the quantities  $J_1^0$ ,  $J_2^0$ ,  $X_1^0$ , and  $X_2^0$ . Then a variation along (12) gives

$$(15) \quad \delta J_1 = L_{11}\delta X_1 + L_{12} \frac{\delta X_2}{\delta X_1} \delta X_1.$$

Multiplying both sides by  $\delta X_1$  and combining with (11) in the form

$$J_1\delta X_1 + J_2\delta X_2 = 0$$

gives

$$(16) \quad \delta J_1\delta X_1 = [(L_{11}L_{22} - L_{12}^2)(\delta X_1)^2(X_2^0)^2]/J_2^0X_2^0.$$

An equivalent result is obtained by exchanging subscripts 1 and 2. Because the matrix of coefficients is positive definite,  $L_{11}L_{22} - L_{12}^2 > 0$ . Furthermore, if  $J_1^0X_1^0 > 0$  and  $J_2^0X_2^0 > 0$  (their sum is necessarily positive, and in the common systems for which the off-diagonal coefficients are relatively small to negligible they are separately positive) then

$$(17) \quad \delta J_1\delta X_1 > 0$$

and

$$(18) \quad \delta J_2\delta X_2 > 0.$$

These state that a change in a thermodynamic force produces a change in its conjugate flux, which is of the same sign. For illustrative purposes we consider two dissipative diffusive fluxes continuously primed in the steady state by non-dissipative convective fluxes. A positive variation will increase the dissipative flux over that required to balance the non-dissipative one and this excess will immediately moderate the perturbing force by decreasing it. Similarly, a negative change in a force will decrease the flux and the convective flow will now be superior to that required to maintain the steady state. It

will accordingly moderate the perturbing force by increasing it. In either case the effect is that demanded by the Le Chatelier principle of moderation. Thus, the configurational state of minimum entropy production rate can be a stable one and therefore define the steady state.

In consideration of heterogeneous systems it becomes apparent why in our description of constraint relation (1) we specified that the constraint lines could wander in time in Cartesian space provided the volume element associated with each constraint line was invariant. This assures that all constraint lines can be chosen so as to be represented as continuous lines on the  $\sigma$ -surface (Fig. 2) even though discontinuities occur in the phenomenological coefficients and singularities occur in the thermodynamic co-ordinates at interfaces. We can accordingly state without further detailed consideration the mathematical equivalence of Euler-Lagrange equations of type (6) and the integral representation (10).

It is probable that our analysis can be generalized to systems of constraints even more complicated than (1) although we have not succeeded in doing this due to the difficulties in formulating the Le Chatelier principle for such systems. Fortunately, the present generality is sufficient for consideration of the systems to be discussed below.

#### *Description of Oscillatory Configurations*

A complex situation may arise in which a constrained system is instantaneously stable against one mode of fluctuation but unstable against another. There will then be a tendency for the system to enter a stationary, oscillatory configuration. Provided the amplitude of the oscillation is small, such a state may still be qualitatively described by the principle of minimum entropy production since it behaves in accord with the Le Chatelier principle, i.e., each local fluctuation is immediately moderated, and the entropy production rate does not change during fluctuation since the fluctuations occur about the extrema in the  $\sigma$ -surface (see Fig. 2). Naturally, the Euler-Lagrange equations, being inherently time independent, will be incapable of describing the instantaneous configuration.

#### AN EXAMPLE OF A STEADY HETEROGENEOUS SYSTEM: DEFINITION OF THE MORPHOLOGICAL FEATURES OF ALLOY SINGLE CRYSTAL GROWTH

In a previous communication (Kirkaldy 1959) the experimental fact was noted from the literature (Tiller and Rutter 1956) that binary alloy single crystals can be grown under true steady state conditions and that for certain boundary conditions a steady cellular morphology develops. We noted, too, that this morphology, as compared to a planar one, can be rationalized by the principle of minimum rate of entropy production since it leads to a corresponding conservation of available energy in the form of solid solute segregation. We were not able, however, to show that the variation between an actual and a flat interface is a valid one with respect to a mathematical formulation of the principle. With our present generalization, applicable to heterogeneous systems with convective flow, it is now possible to do this.

Consider the experimental arrangement shown in Fig. 1 wherein an alloy of concentration  $C_0$  is placed in a temperature gradient defined by two fixed temperature points  $T_1$  and  $T_2$  which bracket the melting range of the alloy. These two points are moved at a fixed velocity  $v$  until a two-dimensional steady state of cellular (non-planar) interface morphology is attained. (We of course assume here that the conditions are chosen so that the internal constraints allow a non-planar morphology.) We then describe the system with forces corresponding to the temperature gradient,  $\mathbf{X}_T$ , the free energy difference for the liquid-solid reaction,  $X_c$ , and two mutually normal components of the one independent diffusional force,  $X_x$  and  $X_y$ .

Now let us describe our system approximately as constrained by fixed forces  $\mathbf{X}_T$  and  $X_c$ , both determined in value by the convective flow rate and the fixed boundary temperature values, and by a constant uniform concentration  $C_0$  entering the system. There remains one degree of freedom in the system arising from the fact that the solute configuration at the exit to the test volume is constrained at the steady state only to the extent that its average value must equal  $C_0$ . The degree of freedom is associated with the statement of this third and final constraint (in a system involving four forces),  $C(T_2) = C_0$ ,  $C(T_1) = C_0$ , in the implicit form

$$(19) \quad F(X_x, X_y, x, y, z) = 0,$$

the functional relationship depending on the specific constraints as well as on the variational path. We choose this path in the interests of tractability so that the extreme varied state is a pseudo-steady one corresponding to a flat interface. This choice also assures the validity of the constraints  $\mathbf{X}_T$ ,  $X_c = \text{constant}$ . We can obtain this path by dropping one of the internal constraints (or imposing a further external one). In particular, we discard any foreknowledge of interfacial correlation processes by regarding  $x'(y, z)$  in (3) as a determinable rather than given function. This makes it possible to determine by standard differential phenomenological procedures the complete configuration of a spectrum of pseudo-steady varied states, and in particular, the configuration corresponding to a flat interface with its uniform output as well as input solute distribution. The procedure used here, in which the time variable is eliminated from the description by the device of passage through pseudo-steady states, is analogous to that used in classical thermodynamics, by means of which the time variable is eliminated by passage through quasi-equilibrium states.

We now incrementally and periodically reimpose the freed internal constraint causing an evolution through the pseudo-steady states towards the final actual steady one. It is assumed that this tractable variational path can be adequately represented by a constraint relation of type (1) when the incremental path between adjacent pseudo-steady states is taken to the limit. Now we know from the constitution diagram and the thermal conditions that the evolution can occur at approximately constant phase volumes, thereby allowing the constraint relation (19) to be formed in such a way that it is always continuous in  $\sigma - X$  space. Thus, in all respects this specification of our real system



corresponds to the ideal one described by relation (12), so the system must have its steady configuration defined by the principle of minimum entropy production (10).

It may reasonably be asked at this point whether a state which is stable against fluctuations along the pseudo path is stable against real fluctuations. We believe this to be the case here, since every pseudo-steady state is a possible instantaneous transient one, i.e., it can always in principle be reached in a real process with fixed constraints by altering the initial conditions.

As noted in our previous communication (Kirkaldy 1959), the change in entropy production rate during such a variational process (initial minus final) can be calculated to the first order as

$$(20) \quad \Delta \int_V \sigma dV = (S_F - S_C) Av,$$

where  $S_F$  and  $S_C$  are the entropies per unit volume at the exit point ( $T_1$ ) for a flat and cellular interface, respectively, and  $A$  is the test area. In other words, the change in entropy production rate is measured by the change in entropy outflow due to convection through the system. Since the cellular morphology leads to a segregated (low entropy) exit state,  $S_C < S_F$ , and  $\Delta$  is positive as required by the variational principle. Accordingly, this somewhat idealized process meets the essential conditions of the phenomenological model and it behaves experimentally in accordance with that model.

We have thus far made no explicit reference to the role of heat conduction and chemical reaction in the process. In fact, it is the flexible coupling between the interfacial reaction position, the thermal field, and the diffusion field via the mass balance and the alloy constitution which affords a decrease in the diffusive entropy production rate without a corresponding increase in the chemical and thermal entropy production rates. This internal flexibility or freedom is to be related to the external degree of freedom and this is in turn defined by the unconfined exit solute distribution. When the constitutional conditions disallow a steady non-planar interface (Walton *et al.* 1955) the variational path considered in this section is also disallowed. There of course remain other variational paths associated with the flexibility of output solute distribution but all of these freely migrate towards a state of uniform output solute distribution as the steady state of minimum rate of entropy production. For a further discussion of the technicalities involved in the definition of the internal constraints, the reader is referred to the Appendix.

This simplest of all heterogeneous dissipative systems is very much more complex phenomenologically than any of the homogeneous ones which have so far been treated analytically so it is not surprising that our understanding of the general features of crystal growth has remained limited. In particular, the following new features are involved: heterogeneous structure, non-dissipative convective flow, and implicit constraints. These latter give rise to the unusual feature of internal behavior that for certain relatively infrequent constraint lines,  $\sigma$  increases towards a maximum at the steady state.



ON THE EXISTENCE OF OSCILLATORY DISSIPATIVE CONFIGURATIONS:  
DENDRITIC GROWTH

In a previous section we noted the possibility that a system may be stable against one mode of fluctuation but unstable against another, giving rise to oscillatory configurations. This is evidently what happens in alloy single crystal growth when the conditions are such as to allow cell lengths of the same order of magnitude as their widths. The barring of new crystal growth faces of low entropy production rate (or high chemical potential) gives rise to the possibility of lateral competing growth directions. For a true steady state it must be possible to state stability conditions against interfacial fluctuations in these lateral directions as well. Presumably, for the more extreme conditions of growth simultaneous stability conditions do not exist so that a breakdown of the steady state of minimum entropy production appears in the form of oscillatory configurations. The result, known as dendritic growth, is shown schematically in Fig. 3.

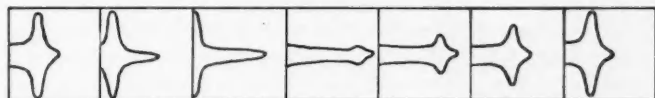


FIG. 3. Representative oscillatory configuration near an advancing dendritic tip.

ON THE NUMERICAL CALCULATION OF THE CONFIGURATION OF  
STEADY HETEROGENEOUS SYSTEMS

The general recipe for the numerical solution of the alloy crystal growth problem is clear enough, though the rendering of the procedure presents a very formidable task indeed. We can assume the quantitative applicability of the principle of minimum entropy production. Initially, it would be advisable to restrict consideration to conditions or materials wherein effects of surface tension on the alloy constitution can be neglected, to growth conditions which produce only slight deviations from planar morphology to assure a true steady state, and to a two-dimensional cell model for arithmetical economy. Then by trial we can in principle determine all those pseudo-steady configurations which satisfy the phenomenological equations, the boundary conditions, and all the explicit internal constraints. Finally, we calculate the volume rate of entropy production for each configuration and select the configuration with the lowest volume rate as the actual one. In line with our assumption of constant thermal and chemical forces we can ignore the contribution of these processes to the entropy production rate. We thus regard the process as entirely diffusion-controlled.

A possible alternative to this final step was suggested in our previous communication (Kirkaldy 1959). We were able to infer an approximate corollary to the principle of minimum entropy production from equation (20) to the effect that the convective flow of entropy out of the system should

also be a minimum. This quantity is undoubtedly easier to calculate than the rate of entropy production within the system and could therefore supply a useful shortcut to the theoretical definition of the steady configuration.

#### SUMMARY

The problem of mathematical description of a constrained, convective, heterogeneous dissipative system is generically identical with that for a non-convective homogeneous system. Unique solutions demand the removal of any degree of freedom in the system by the specification of steady state conditions. These are patently obtained via the Le Chatelier principle of moderation. We have demonstrated that for certain interesting convective heterogeneous systems, as with non-convective homogeneous systems, the variational principle of minimum integrated rate of entropy production generates the Le Chatelier principle via the Euler-Lagrange equations derived therefrom. While the differential description of the system is implicitly complete, it is not capable of producing a numerical description of the configuration. This can only be obtained by trial using procedures based on the variational description of the system.

#### ACKNOWLEDGMENTS

The author is indebted to Dr. W. A. Tiller of Westinghouse Research Laboratories, Pittsburgh, and Dr. J. R. Schrieffer of the University of Illinois, Urbana, for helpful discussions, to Dr. R. R. Haering of McMaster University for his comments on the manuscript, and particularly to the referee for pointing out certain errors of commission and omission. Financial assistance in support of the associated experimental program by the National Research Council of Canada, and the American Iron and Steel Institute is gratefully acknowledged.

#### REFERENCES

- BROWN, W. B. 1958. *Trans. Faraday Soc.* **54**, 772.  
 DE GROOT, S. R. 1952a. *Thermodynamics of irreversible processes* (North-Holland Publishing Co., Amsterdam).  
 ——— 1952b. *Thermodynamics of irreversible processes* (North-Holland Publishing Co., Amsterdam), p. 196.  
 ——— 1952c. *Thermodynamics of irreversible processes* (North-Holland Publishing Co., Amsterdam), p. 197.  
 DENBIGH, K. G. 1951. *The thermodynamics of the steady state* (Methuen and Co. Ltd., London).  
 KIMBALL, W. S. 1952. *Calculus of variations* (Butterworth Scientific Publications, London), p. 278.  
 KIRKALDY, J. S. 1959. *Can. J. Phys.* **37**, 739.  
 PRIGOGINE, I. 1947. *Etude thermodynamique des phénomènes irréversibles* (Liege, Ed., Desoer).  
 ——— 1955. *Introduction to thermodynamics of irreversible processes* (Charles C. Thomas, Publisher, Springfield, Ill.).  
 TILLER, W. A. 1959. *Can. J. Phys.* **37**, 1204.  
 TILLER, W. A. and RUTTER, J. W. 1956. *Can. J. Phys.* **34**, 96.  
 WALTON, D., TILLER, W. A., RUTTER, J. W., and WINEGARD, W. C. 1955. *Trans. Am. Inst. Mining Met. Engrs.* **203**, 1023.

## APPENDIX

## SOME COMMENTS ON THE DEFINITION OF THE INTERNAL CONSTRAINTS

In our earlier paper (Kirkaldy 1959) we noted that a condition of maximum segregation (minimum entropy outflow) consistent with both the external and internal constraints could be regarded as an approximate corollary to the principle of minimum rate of entropy production. Since the amount of segregation will depend monotonically on the length of the cells, the cells should be as long as possible consistent with the constraints. The primary internal constraint noted was the constitutional requirement of substantial equilibrium at all points of the interface. Since the sides of a cell lie along a temperature gradient we inferred that the length of the cell is limited only by the condition that the actual temperature gradient in the growth direction at a cell tip should not exceed the gradient of the liquidus temperature distribution as computed from the solute distribution, and for maximum length should be equal to it. In the terminology of the Chalmer's school this is equivalent to saying that no "constitutional supercooling" exists in the liquid at the cell tip. Tiller (1959) has criticized this inference on the grounds that our proposed condition of maximum length, limited only by constitutional considerations, may lead to a violation of another internal constraint, the solute balance. We acknowledge this criticism with the following comments:

1. There are certain conditions near the critical ones for onset of cellular growth wherein both constraints can be simultaneously satisfied and for these our original statement is probably valid. We suspect that this region corresponds to the observed steady state conditions where the cells remain completely featureless.

2. For rather more extreme conditions the solute balance limits the cell length in such a way that constitutional supercooling "relative to a flat interface" is instituted. However, for these more extreme conditions there is a tendency for the cells to develop teats with a very small radius of curvature. It is conceivable that the surface tension due to this curvature, which Tiller mentions, is sufficient to alter the constitution of the tip so as to increase the segregation (or lower the entropy production rate) and at the same time maintain a condition of no constitutional supercooling "relative to the curved interface". Thus, if we correctly regard the term "constitutional supercooling" as a relative quantity then our original statement could still be valid for all steady conditions.

# THE THERMODYNAMIC DESCRIPTION OF HETEROGENEOUS DISSIPATIVE SYSTEMS BY VARIATIONAL METHODS

## II. A VARIATIONAL PRINCIPLE APPLICABLE TO NON-STATIONARY (UNCONSTRAINED) HETEROGENEOUS DISSIPATIVE SYSTEMS<sup>1</sup>

J. S. KIRKALDY

### ABSTRACT

The variational principle

$$\delta \int_{\text{volume}} [2 \sum_i J_i X_i - \sum_{i,k} R_{ik} J_i J_k] dV = 0$$

involving the independent thermodynamic fluxes  $J_i$  and forces  $X_i$  and subject to the non-holonomic constraint,  $X_i = \text{constant}$ , gives an expression for the integral behavior of an unconstrained heterogeneous conduction-diffusion-reaction-viscous flow process. The validity of this expression can be checked by performing the variation with respect to the forces to obtain as Euler-Lagrange equations the phenomenological equations,

$$X_i = \sum_k R_{ik} J_k.$$

This principle allows the unique mathematical specification of certain non-stationary systems which are not easily amenable to differential analysis.

As an example, it is demonstrated that the principle generates an approximate expression for the steady growth velocity,  $v$ , of an isothermal segregation reaction in terms of the degree of advancement of the reaction,  $\phi$ , and its derivative with respect to  $v$ ,

$$v = \phi(v)/\phi'(v).$$

### INTRODUCTION

In our accompanying paper I (Kirkaldy 1960) we discussed the thermodynamic behavior of a constrained system with an internal phase interface. This is one of a class of dynamic systems which does not easily submit to unique phenomenological description in terms of differential analysis alone. The difficulty in that case lay in two directions: firstly, the constraints were given only in an integral form and so for application to the phenomenological description these were necessarily regarded as implicit, and secondly, certain internal constraints, relating to co-operative phenomena at the interface were not expressible in phenomenological terms. As we demonstrated, it was still possible to establish an optimal, integral representation of this system capable of generating a unique solution of a steady state configuration by trial procedures.

In the present contribution we consider the problem of the instantaneous definition of the configuration of a heterogeneous system which is unconstrained. Such a system will evolve towards equilibrium and will be inherently non-steady during this evolution. The previous success of an integral representation in overcoming the mathematical difficulty associated with discontinuous thermodynamic co-ordinates suggests that similar methods are applicable here.

<sup>1</sup>Manuscript received March 7, 1960.

Contribution from the Department of Metallurgy and Metallurgical Engineering, McMaster University, Hamilton, Ont.

Onsager (1945-46) anticipated the variational approach to irreversible thermodynamics by describing isothermal diffusion in a materially closed homogeneous system by means of the variational principle,

$$(1) \quad \delta \left[ -\frac{dF}{dt} - \frac{1}{2} \int_v T \sigma dV \right] = 0, \text{ maximum.}$$

In this,  $F$  is the total Gibbs' free energy,  $T$  is the absolute temperature, and  $\sigma$  is the entropy production rate per unit volume. Onsager showed that  $-dF/dt$  can be expressed as an integral over a bilinear expansion in the thermodynamics forces and fluxes. When the entropy production rate is then expressed as a symmetric, positive definite, quadratic function in the fluxes, the variational principle is confirmed, on variation with respect to the fluxes with fixed forces, by generating as Euler-Lagrange equations the experimentally observed inverse phenomenological equations,

$$(2) \quad X_i = \sum_k R_{ik} J_k.$$

While (1) was stated originally as a generalization having academic value only as applied to homogeneous systems, the writer has suggested that such a statement may serve a practical purpose in the description of heterogeneous reaction-diffusion systems (Kirkaldy 1959a) and general heterogeneous dissipative systems (Kirkaldy 1959b). In the present contribution we carry out the mathematical procedure which generalizes Onsager's variational principle so as to apply to unconstrained conduction-diffusion-reaction-viscous flow systems and consider the possible methods of application to heterogeneous systems.

With foreknowledge of Onsager's result and in anticipation of our analysis we can state that a variational description of such a system will have the following (not necessarily mutually exclusive) properties:

1. It cannot violate the local phenomenological relations. In fact, the principle constructed generates these relations as Euler-Lagrange equations.
2. Since for simplicity we desire a variational formulation which does not explicitly contain the time, the variational integral must be a volume integral and the variation must be carried out at fixed time.
3. The requirement that the time be fixed during variation bespeaks the impossibility of a variation which connects possible dynamic paths, and implies that dynamic constraints corresponding to the fixed time must be imposed on the integrand through the thermodynamic co-ordinates in opposition to certain internal constraints.
4. The principle must manifest in some way the high dynamic efficiency implied by the principle of minimum entropy production for steady state systems (paper I). This suggests that the variational integrand should depend in a direct way on the rate of entropy production.

#### THE GENERAL VARIATIONAL PRINCIPLE

Consider a dissipative (conduction-diffusion-reaction-viscous flow) thermodynamic system which can be described by  $n$  independent forces  $X_i$  and

conjugate fluxes  $J_i$  and let the rate of irreversible entropy production per unit volume be represented as a symmetric, positive definite, quadratic expression in the independent fluxes,

$$(3) \quad \sigma = \sum_{i,k} R_{ik} J_i J_k$$

where the connecting phenomenological equations (2) generate the alternative bilinear form

$$(4) \quad \sigma = \sum_i J_i X_i.$$

In (3) the coefficients  $R_{ik}$  are independent of the fluxes but may be regarded as functions of the Cartesian co-ordinates. For example, in surface reaction we can introduce coefficients of the form

$$(5) \quad R_{ik} = \mathcal{R}_{ik} \delta[x - x'(y, z)]$$

which explicitly recognize that a surface process is a singular volumetric one. Similar singularities would have to be allowed for if the diffusive process involves interfacial as well as volume diffusion.

Now let us introduce the function

$$(6) \quad \xi = 2 \sum_i J_i X_i - \sum_{i,k} R_{ik} J_i J_k$$

and impose the non-holonomic constraint  $\delta X_i = 0$ , or

$$(7) \quad X_i = \text{constant}.$$

$X_i$  is to be regarded here as a type of Lagrange multiplier to be explicitly evaluated later. We can then show that provided  $X_i$  takes the constant value equal to the  $i$ th force at the extremum,

$$(8) \quad \delta \int_V \xi dV = 0,$$

generates the phenomenological equations (2) via the Euler-Lagrange equations\*

$$(9) \quad (d\xi)_{X_i} = 0,$$

variation being carried out along any line in the  $\xi$ -surface. The maximal nature of the extremum can be demonstrated as a consequence of the positive definite character of  $[R_{ik}]$  by showing that

$$(10) \quad (d^2\xi)_{X_i} = \text{negative}.$$

Even more simply we can recognize the nature of this extremum by noting that  $\xi$  passes through the origin and, as well, reaches its extremum on the line  $\xi = \sigma$ , the latter being positive definite (see Fig. 1).

Keep in mind that the constraint (7) connects dynamic paths all but one of which are impossible ones since these paths can only be maintained by

\*This construction is formally related to Hamilton's principle for a dissipative, non-holonomic mechanical system (Whittaker 1944).

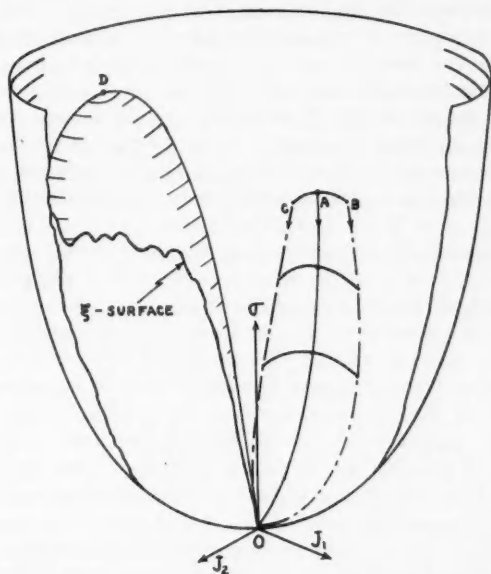


FIG. 1. Schematic two-flux  $\sigma$ -surface indicating representative unconstrained dynamic paths and a representative intersecting  $\xi$ -surface.

violating certain internal constraints. However, since a valid variational path for (8) is any line which passes through the extremum, we can impose a further condition on the path to the effect that it is restricted to the  $\sigma$ -surface, i.e., to a particular line passing through the extremum,

$$(11) \quad \xi = \sigma.$$

Thus, we can assign a definite rate of entropy production to a varied state even though such a state can be maintained only by violating the internal constraints. This restriction will be found useful in our later applications.

The geometric representation is given schematically for a two-flux system in Fig. 1. Point  $D$  represents the maximum along a typical intersection of a  $\sigma$ - and  $\xi$ -surface. Similarly, point  $A$  for another volume element lies at the maximum of a line  $BC$  describing the intersection of the  $\sigma$ -(equation 3) and  $\xi$ -surfaces. As a result of (11) this maximum in  $\xi$  and its volume integral might be thought to contradict the principle of minimum entropy production applicable to steady systems. This, however, is not the case, since the location of the extremum of the constrained function  $\xi$  is determined mainly by the high curvature of its unconstrained second term,  $-\sigma$ . Thus, we may say qualitatively that the rate of entropy production is a minimum consistent with a dynamic constraint which assures a non-zero transformation rate. This point will be clarified when we examine a specific system in a later section.

Alternative evidence for the existence of an over-all high dynamic efficiency in unsteady systems can be inferred from Fig. 1. Consider two dynamic paths  $AO$  and  $BO$ , the first being the actual one and the second being a varied one which violates the internal constraints. At any time  $t$  the actual path has a greater  $\sigma$  than the varied one. If we impose on the system the requirement that the total dissipation of available energy be the same for all paths then the degree of advancement of a reaction at time  $t$  is always greater for the varied path than for the actual one and the latter is therefore the more efficient dynamically.

A sufficient condition for application of the foregoing to heterogeneous systems is that that part of the total volume of each phase which contains dissipative processes can be assumed invariant along the over-all variational path. This makes it possible for the variational path of a single volume element associated with (9) to wander in Cartesian space in such a way that no discontinuities occur. We are not sure, however, that this is a necessary condition for the validity of the variational integral. For qualitative purposes we would not hesitate to apply the principle to systems which can meet all the other conditions of the formulation, even though they do not satisfy this one. A second condition is that it should be possible to choose the thermodynamic fluxes so that they are independent in the sense required by the variational principle (8). The coupling between reaction and diffusion processes in general denies this independence. However, since it is necessary for the application of the variational principle to release some internal constraint, we shall choose that constraint associated with the coupling between the two processes and in this way assure the independence required by the phenomenological scheme.

In line with our observations of constrained systems (paper I) we expect that competing instabilities in unconstrained systems will in certain circumstances lead to oscillatory configurations. Dendritic growth of solidifying alloy castings or of sublimating snow crystals are examples of such behavior.

It should finally be appreciated that for an unconstrained system the procedure outlined is perfectly symmetric in the fluxes and forces. If the rate of entropy production is transposed and written as a symmetric, positive definite quadratic function in the forces instead of the fluxes,

$$(12) \quad \sigma = \sum_{i,k} L_{ik} X_i X_k,$$

then we can obtain from (8) with  $J_i = \text{constant}$ , the inverse phenomenological equations,

$$(13) \quad J_i = \sum_k L_{ik} X_k.$$

#### AN EXAMPLE OF A NON-STEADY DISSIPATIVE SYSTEM: THE APPROXIMATE MATHEMATICAL DESCRIPTION OF THE MORPHOLOGICAL DEVELOPMENT OF THE ISOTHERMAL SEGREGATION REACTION

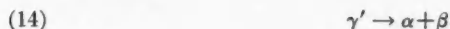
A very common type of experimental non-steady system is one which tends from an initial well-defined state towards a final equilibrium state. Such a



system has no external constraints and if it contains discontinuities in the thermodynamic co-ordinates it presents a special problem for the definition of the process path in the  $\sigma$ -surface.

A complete definition is, of course, impossible without reference to a time-dependent description and to fluctuation phenomena. However, if we have obtained an array of "possible" time-dependent paths through trial calculations and (or) observations, a method presents itself for the selection of the actual path from the "possible" paths based on the variational principle (8). A "possible" path in the sense used here is one which is not subject to rigid chemical coupling at the interface and is therefore free from spontaneous morphologic transients.

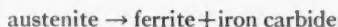
Consider as the simplest examples those solid state reactions described symbolically by



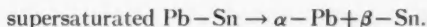
and



which occur substantially isothermally and at a constant rate  $v$  on quenching from a temperature at which  $\gamma$  or  $\alpha$  is stable to some supersaturated state. These develop by the mode of linear growth of the new phases into the old. With Cahn (1959) we characterize these as segregation reactions and take as examples the eutectoid reaction in steel (the pearlite reaction)

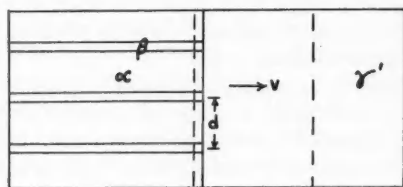


and the cellular segregation reaction in lead-tin

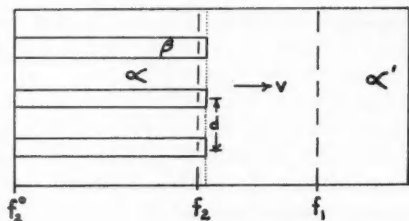


The idealized growth configurations are shown in Fig. 2. For these two reactions the phase cross sections are platelike, but for the purposes of the following analysis they can have any cross section.

Although such systems are inherently non-stationary since we know only the initial and the final equilibrium conditions, they ultimately attain a steady configuration relative to the moving interface. Ignoring the thermal constraints since the substantially zero thermal forces do not contribute to the dissipation, we can regard the systems as being externally unconstrained during transformation. The theoretical problem which has been raised by previous workers (Zener 1946; Turnbull 1955; Hillert 1957; Cahn 1959) is: Given the general morphological layout of Fig. 2 how can one predict the cell spacing  $d$  and the uniform growth velocity  $v$ ? We know, in principle, that if  $v$  is given we can calculate  $d$  using the phenomenological equations and standard differential procedures, and vice versa, but, in fact, neither is given. What is the source of this apparent internal degree of freedom and how can it be removed? In answer to the first, the apparent degree of freedom arises from our lack of knowledge of the modes of coupling between the diffusion and interfacial reaction processes. In effect, by neglecting this, we release



(a)



(b)

FIG. 2. Schematic configurations of segregation reactions. (a) The eutectoid reaction, (b) the cellular segregation reaction. Symbols are defined in text.

an internal constraint. In order that a description which neglects this be made unique, a suitable alternative constraint must be used. The variational formalism introduced above appears to meet the requirements of the problem in this respect and to provide an optimal condition for specifying the actual configuration. The above authors recognized the need for such an optimizing principle but their substitution of an *ad hoc* statement for a principle properly based on the variational calculus makes their results extremely suspect.

The procedure is to first obtain a spectrum of "possible" paths in the  $\sigma$ -surface. A suitable set can be obtained by dropping the internal constraint associated with interfacial co-ordination and calculating the complete configuration associated with a set of values of the growth velocity  $v$ . Then, tentatively assuming each value of  $v$  in turn as the optimum one, we evaluate a series of volume integrals about the trial configuration subject to the appropriate substitute constraint ( $X_i = \text{constant}$ ). That velocity for which the trial volume integral is a maximum among the varied integrals,  $v_m$ , is to be regarded as the actual one. Note that in this procedure a complete variational calculation must be performed for each of the "possible" trial configurations. This very uneconomic situation is a result of the fact that the variation is a constrained one and the specification of the constraint is implicit, i.e., it depends on the values of the thermodynamic co-ordinates at the actual extremum.

As an approximate alternative method the following analytic procedure may be used for the configurations of Fig. 2. We again use the experimentally

observed fact that the growth velocity is constant for appreciable periods of time and we assume that the rate of desegregation reaction behind the advancing interface is negligible. Even if the latter is a poor approximation in itself, the assumption should lead to little error in the determination of the interface configuration since this will be most sensitive to what occurs at and ahead of the interface. (The fact that the velocity appears constant is the best justification of the approximation.) These approximate specifications allow the choice of constant phase volumes during variation and assure a tractable form for the analysis.

We can now write the variational integral as

$$(16) \quad \int \xi dV = 2Av(f_1 - f_2) - Av\phi(v)(f_1 - f_2^0)$$

where the integrand

$$(17) \quad \xi = [2\sum_i J_i X_i - \sum_{i,k} R_{ik} J_i J_k]_{X_i = \text{constant}},$$

where the degree of advancement of the reaction  $\phi(v)$  is defined by

$$(18) \quad f_1 - f_2 = \phi(v)(f_1 - f_2^0),$$

and where  $f_1$ ,  $f_2$ , and  $f_2^0$  represent the mean available energy per unit volume in the unreacted material, in the material immediately behind the interface, and in the final equilibrium material, respectively. If the system can be assumed to be isobaric as well as isothermal these quantities represent Gibbs' free energies. In characterizing the variation by a single parameter  $v$  we effectively regard both  $\xi$  and its volume integral as explicit functions of  $v$  and we have expressed this explicit behavior by (16). Furthermore, the constraint  $X_i = \text{constant}$  in (17) finds a uniquely equivalent statement in (16) to the effect that  $f_1 - f_2$  in the first term is constrained to a constant value. This is because the total driving potential difference  $f_1 - f_2$ , must be constant if the incremental differences measured by the  $X$ 's are constant. The second term,  $-\int \sigma dV$ , involves no constraint and so its  $v$ -dependence is stated in full,  $f_1 - f_2^0$  by definition being a constant. The variational principle therefore becomes

$$(19) \quad (d/dv)[2v\phi(v_m) - v\phi(v)] = 0,$$

the implicit nature of the constraint still being apparent in the constant  $\phi(v_m)$ . This leads to a transcendental algebraic equation

$$(20) \quad v_m = \phi(v_m)/\phi'(v_m),$$

which in general must be solved numerically.

It should finally be remarked that the tractability of the suggested procedures is explicitly dependent on our ability to represent an inherently non-steady process by an internally steady configuration since this eliminates the necessity of matching trial solutions on the time scale.

While the problem is now completely solved in principle, the calculation is still a long way from producing numerical predictions. An explicit statement

of  $\phi(v)$  requires a precise knowledge of the reaction mechanism and all the phenomenological coefficients pertaining thereto. In discussing the pearlite reaction, for example, reference in the literature has been made to the influence of volume diffusion, surface reaction, surface diffusion, and plastic deformation. If all of these are significant, the numerical problem is indeed discouraging, particularly if viscous flow is involved. It is, in fact, doubtful that a linear phenomenological scheme, specifically applicable to Poiseuille flow of a Newtonian fluid, is at all suitable for the description of flow in solids.

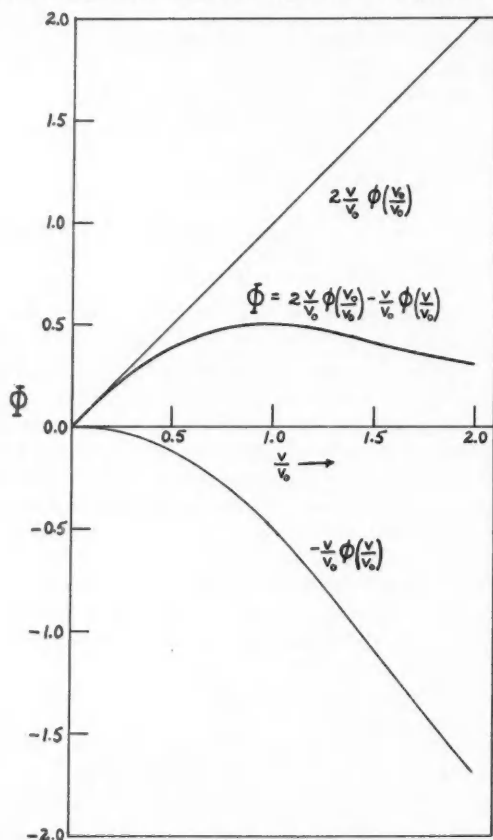


FIG. 3. Graphical representation of dissipation functions used in the description of unconstrained dissipative systems.

While the present outlook for useful numerical predictions is not optimistic, the general scheme still provides us with insight into the qualitative behavior of complex dissipative systems. For example, we can construct the degree of advancement function

$$(21) \quad \phi(v) = 1 - \exp \left\{ - \left( \frac{v}{v_0} \right)^{\frac{1}{\ln 2}} \ln 2 \right\}$$

in such a way that the integral rate of entropy production properly goes to zero as  $v \rightarrow 0$ , and so that (19) operating upon it specifies a maximum at  $v_m = v_0$ . If we then plot a function  $\Phi$  which is proportional to the variational integral (16)

$$(22) \quad \Phi = 2 \frac{v}{v_0} \phi\left(\frac{v_0}{v_0}\right) - \frac{v}{v_0} \phi\left(\frac{v}{v_0}\right)$$

as in Fig. 3, it becomes clear how each of the terms contributes to the location of the extremum and, in particular, how the tendency towards a minimum in the entropy production rate (represented by the second term in (22)) is manifested in a non-steady process. This supports our earlier contention that  $\sigma$  is a minimum consistent with a dynamic constraint which assures a non-zero transformation rate.

#### SUMMARY

The central aim of these companion papers has been to demonstrate that there exists a unique phenomenological description of the growth features of certain heterogeneous dissipative systems. This means that morphological features are essentially thermodynamic in nature and can be so described uniquely if the thermodynamic data is detailed enough. The basic difficulty in such a description arises from the fact that the surface processes are coupled to the diffusion field in a manner which cannot be suitably related to a conventional phenomenological description. It is demonstrated that in those systems which can be assumed to have a "steady" internal configuration (these may be inherently steady or unsteady systems) the coupling constraint can be replaced by a mathematical constraint which is suitable to a conventional phenomenological formalism. This makes it possible to uniquely define the system in terms of variational principles. These can be used to define the configuration of the system by trial methods.

#### ACKNOWLEDGMENTS

The author is indebted to Dr. J. W. Cahn of General Electric Research Laboratories, Schenectady, N.Y., and to Professor M. A. Preston of McMaster University for helpful discussions, and to Dr. R. R. Haering of McMaster University for his comments on the completed manuscript. Financial assistance in support of the associated experimental program by the National Research Council of Canada and the American Iron and Steel Institute is gratefully acknowledged.

#### REFERENCES

- CAHN, J. W. 1959. *Acta Met.* **7**, 18.
- HILLERT, M. 1957. *Jernkontorets Ann.* **141**, 757.
- KIRKALDY, J. S. 1959a. *Trans. Am. Soc. Metals*. Discussion to paper by J. B. Clark and F. N. Rhines, p. 220.
- 1959b. *Proc. International Symposium on the Physical Chemistry of Process Metallurgy*. In press.
- 1960. *Can. J. Phys.* **38**. This issue.
- ONSAGER, L. 1945-46. *Ann. N. Y. Acad. Sci.* **46**, 241.
- TURNBULL, D. 1955. *Acta. Met.* **3**, 55.
- WHITTAKER, E. T. 1944. *A treatise on analytical dynamics* (Dover Publications, New York) p. 248.
- ZENER, C. 1946. *Trans. A.I.M.E.* **167**, 550.

## FURTHER OBSERVATIONS OF THE HORIZONTAL MOVEMENTS OF AURORA<sup>1</sup>

J. A. KIM<sup>2</sup> AND B. W. CURRIE

### ABSTRACT

A previous paper showed that systematic motions of aurora parallel and normal to the geomagnetic meridians could be deduced from successions of all-sky camera photographs. More reliable deductions which are based on a much larger number of observations, including some from a station inside the auroral zone, are reported in this paper. Both southward and northward speeds increase with geomagnetic latitude to the auroral zone where they become constant, or even decrease slightly, before continuing to increase inside the auroral zone. A seasonal variation of north-south speeds does not appear to exist. A diurnal variation of north-south speeds, dependent on geomagnetic latitude, is evident, the speeds decreasing to a minimum during the morning hours to the south of the auroral zone and increasing to a maximum during the same hours to the north of the zone. Large positive correlations exist between north-south speeds and departures of the magnetic field intensity from normal. East-west speeds either decrease or remain constant with increasing geomagnetic latitude. A significant seasonal variation of east-west speeds is not evident. Eastward speeds are at a maximum between 03 and 04 hours L.M.T., and are associated with negative magnetic bays; westward speeds, between 21 and 22 hours L.M.T., and are associated with positive magnetic bays.

### INTRODUCTION

A previous paper by Kim and Currie (1958) showed that systematic motions of aurora parallel and normal to the geomagnetic meridians could be deduced from successions of all-sky camera photographs of aurora. These motions did not show evidence of a component due to the earth's rotation relative to fixed excitation patterns in space. The distribution and magnitude of the speed of the auroral structures parallel and normal to the geomagnetic meridians were substantially the same as for non-luminous irregularities observed by radio methods. The speeds increased with geomagnetic activity, particularly in an east-west direction. The number of observations were too few to make significant deductions concerning the geographical, diurnal, and seasonal variations of the motions. The geographical coverage has been improved by the addition of data from Aklavik which is located on the northern edge of the zone of maximum auroral occurrence. The number of observations on north-south motions for different arcs or bands has been increased from 77 to 358 for Saskatoon, from 587 to 974 for Flin Flon, and from 502 to 642 for Uranium City. The data from Aklavik yielded 237 observations. Measurements on east-west motions, made only for Flin Flon, have been increased from 90 to 250.

The reader should consult the afore-mentioned paper for the procedures used in deducing the motions. As before, each speed is applicable to a distinct arc or band for north-south motions, and is based on not less than four successive

<sup>1</sup>Manuscript received July 11, 1960.

Contribution from the Institute of Upper Atmospheric Physics, University of Saskatchewan, Saskatoon, Sask.

<sup>2</sup>Present address: Physics Department, University of Idaho, Moscow, Idaho, U.S.A.

photographs of the auroral structure taken at 1-minute intervals. In the majority of cases, they are based on eight or more successive photographs. Detection of east-west motions depend on convolutions in the auroral structures. A recognizable convolution seldom persists for more than a few minutes. Hence, the east-west speeds are less reliable than the north-south speeds. Only east-west speeds deduced from three or more successive photographs taken at 1-minute intervals are considered.

## RESULTS

### 1. *Motions Parallel to the Geomagnetic Meridians*

Table I gives the geographic latitude and longitude ( $\phi$ ,  $\lambda$ ), the geomagnetic latitude and longitude ( $\Phi$ ,  $\Lambda$ ), and the angle  $\Psi$  between the geographic and the geomagnetic meridian for each station. From the geomagnetic latitudes,

TABLE I

Geographic latitude and longitude ( $\phi$ ,  $\lambda$ ), geomagnetic latitude and longitude ( $\Phi$ ,  $\Lambda$ ), and the angle  $\Psi$  between the geographic and geographic meridians for Saskatoon, Flin Flon, Uranium City, and Aklavik

Station	$\phi$ (°N.)	$\lambda$ (°W.)	$\Phi$ (°N.)	$\Lambda$ (°W.)	$\Psi$ (°)
Saskatoon	52.1	106.6	60.5	48.9	13.8
Flin Flon	54.8	101.8	63.8	43.5	13.6
Uranium City	59.6	108.8	67.7	55.6	19.1
Aklavik	68.2	135.0	70.9	95.0	31.8

Aklavik should be located inside the zone of maximum auroral occurrence for the Northern Hemisphere. Actually, the percentage occurrence of aurora at Aklavik during the months of observation (December, 1957, January and February, 1958) was slightly greater than for the corresponding months at Uranium City, suggesting that the auroral zone for this period was close to the position determined by Fritz (1881).

The much larger number of observations from Saskatoon, Flin Flon, and Uranium City on the motions of arcs and bands parallel to the geomagnetic meridians has not altered the general conclusions mentioned in the first paper. These were the approximate constancy of the speed of an auroral structure when measured for successive 4-minute intervals of time, the infrequent occurrence of arcs or bands that are stationary although particular ones sometimes reverse their direction of motion, the greater percentage of structures showing southward motions (about two-thirds for each of the three stations to the south of the auroral zone), and the frequency distribution of speeds. The average speeds deduced from the larger number of observations are 8.2, 9.3, and 11.7 km/min respectively for southward motion at Saskatoon, Flin Flon, and Uranium City, and 9.5, 10.0, and 10.6 km/min for northward motion. Applying the usual statistical tests, the difference in southward and northward speeds for each of the stations is not significant. This is contrary to the conclusions from the smaller number of observations.

The Aklavik observations showed almost equal occurrences of southward and northward motions (125 to 112) and flatter distributions of the occurrence of various speeds. The first reflects the common impression of observers inside the auroral zone that displays form along the auroral zone and move northward as they develop; the second led to larger average speeds, 12.1 km/min for southward motion and 13.3 km/min for northward motion.

The extent to which the higher speeds for Aklavik are due to a variation with the solar cycle is uncertain. The Aklavik values were deduced from photographs taken from one to two years later than for the other stations. The speeds of southward and northward motions for the four stations were tabulated according to geomagnetic latitude. The latitude assigned to each speed was that above which the display was located at the time corresponding to the middle of the interval for the speed determination. Figure 1 shows the

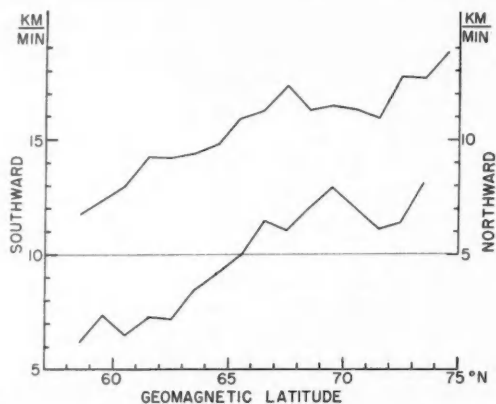


FIG. 1. Variation of average speeds, southward and northward respectively, with geomagnetic latitude. The upper curve is for northward speeds; the lower, for southward speeds.

average speeds for each one-degree interval of geomagnetic latitude, plotted against the corresponding geomagnetic latitude. Both the southward and northward speeds increase fairly regularly with increasing geomagnetic latitude until the center of the auroral zone (approximately  $\Phi = 68^\circ$  N.) is reached. Northward from the center of the auroral zone the speeds become constant, or even decrease slightly, before continuing to increase. The usual statistical tests showed that the latitudinal variation of speeds to the south and through the auroral zone is real. More observations are required to establish the latitudinal variation of speeds inside the auroral zone.

The seasonal variations of the southward and northward speeds parallel to the geomagnetic meridians were examined using the Saskatoon and Flin Flon observations for 1956. Uranium City and Aklavik observations were not used, since photographs cannot be taken at either of these stations for several months during the summer because of daylight. The average speeds for each



month showed an irregular variation. If a seasonal variation exists, many hundreds of observations will be required to establish it.

The diurnal variations of the southward and northward speeds parallel to the geomagnetic meridians were examined by grouping speeds according to each hour of local time. A significant variation that could be attributed to direction of motion (southward or northward) was not apparent. For this reason, the observed speeds during each hourly interval at each station were averaged without regard to direction of motion. Values of the average hourly speeds are given in Table II. An examination of the tabulated speeds suggests that speeds parallel to the geomagnetic meridians decrease to a minimum during the morning hours for stations outside the auroral zone (Saskatoon

TABLE II

Average speeds parallel to the geomagnetic meridians according to local mean times at each station. The quantity in parentheses after each speed is the number of occurrences on which the average is based

L.M.T. (hours)	Average speeds (km/min)			
	Saskatoon	Flin Flon	Uranium City	Aklavik
17-18			11.1 (3)	
18-19	18.5 (2)	8.1 (11)	7.7 (18)	10.1 (9)
19-20	10.9 (6)	10.6 (20)	9.5 (35)	7.0 (19)
20-21	11.0 (11)	11.1 (50)	10.8 (78)	10.0 (26)
21-22	10.7 (31)	10.3 (82)	10.9 (78)	11.6 (26)
22-23	9.5 (49)	11.1 (158)	11.5 (69)	11.9 (32)
23-24	9.3 (64)	10.2 (178)	11.9 (80)	13.0 (32)
00-01	8.9 (45)	9.8 (178)	11.3 (85)	13.8 (29)
01-02	8.0 (54)	9.0 (146)	10.8 (74)	14.7 (14)
02-03	7.7 (52)	8.6 (64)	11.8 (58)	15.0 (15)
03-04	6.8 (23)	7.9 (45)	10.8 (48)	13.7 (6)
04-05	10.7 (10)	10.1 (37)	10.9 (15)	13.0 (14)
05-06		12.5 (12)		13.8 (6)

and Flin Flon), are relatively constant throughout the night for stations along the auroral zone (Uranium City), and increase to a maximum during the morning hours for stations inside the auroral zone (Aklavik). The average speeds for the early evening and late morning hours are less reliable than for the middle portion of the night, because of the much smaller number of observations available for computing them; but they are consistent in showing increased speeds for all stations at these times.

A positive correlation exists between north-south speeds and the simultaneously occurring departures of the magnetic field intensity from normal. The magnetic deviation,  $\Delta F$ , was computed from the values of  $\Delta X$ ,  $\Delta Y$ , and  $\Delta Z$  ( $X$ ,  $Y$ , and  $Z$  are the components southward, eastward, and downward of the earth's magnetic field), for the time at which the motion occurred. Speeds were grouped according to 1 km/min intervals and averaged for each group. The corresponding magnetic deviations were also averaged. The correlation coefficients for Saskatoon are  $+0.79$  for southward motion and  $+0.84$  for northward motion. In the correlation, the reaction between the speed and the magnetic disturbance may not be mutual in that one factor

(say the speed) is the causative agent which produces by a change in magnitude some measurable response in the second factor (say the magnetic disturbance). Regression equations giving the relationships between the variates, assuming first one and then the other as the causative factors, were found.

The equations for Saskatoon and Flin Flon are as given below. The symbol  $S$  is used in the equations to designate speed in km/min;  $\Delta F$ , magnetic departure from normal in gammas.

Northward motion:

$$\left. \begin{array}{l} S = 0.12\Delta F - 5.0 \\ \Delta F = 6.1S + 62.5 \\ S = 0.14\Delta F - 6.8 \\ \Delta F = 5.5S + 65.7 \end{array} \right\} \begin{array}{l} \text{Saskatoon} \\ \text{Flin Flon} \end{array}$$

Southward motion:

$$\left. \begin{array}{l} S = 0.09\Delta F - 3.0 \\ \Delta F = 7.0S + 67.3 \\ S = 0.11\Delta F + 5.6 \\ \Delta F = 7.9S + 61.7 \end{array} \right\} \begin{array}{l} \text{Saskatoon} \\ \text{Flin Flon} \end{array}$$

It should be noted that the equations for the magnetic departures give values between 60 and 70 gammas for auroral structure without north-south motion. While arcs and bands without north-south motions are infrequent in this region, they do occur, as pointed out by Weaver and Skinner (1960). It is also possible that the magnetic departures corresponding to zero speeds are caused by east-west motions which cannot be observed except under more disturbed conditions of the magnetic field.

## 2. Motions Normal to the Geomagnetic Meridians

Studies of motions normal to the geomagnetic meridians were made for Flin Flon (as mentioned in the Introduction). The increase in the number of observations (from 90 to 250) made possible qualitative deductions concerning the geographical and seasonal variations of the speeds, and more reliable deductions concerning the diurnal variations of the speed and the type of magnetic bay associated with the direction of motion.

The average speeds are much greater than for the north-south speeds, a fact that might be anticipated from the high magnetic activity that is associated with auroral structures suitable for detecting the motion. The average speeds were 31.7 and 31.4 km/min respectively for eastward and westward motions. These values are substantially the same (28.9 and 28.7 km/min) as reported in the first paper. Again, the increase in speeds may be associated with increased solar activity, the additional measurements being for 1957.

Figure 2 shows the variation in eastward and westward speeds with geomagnetic latitude. Speeds were grouped according to one-degree intervals of latitude over which the aurora was located at the time of the photograph, and then averaged. The eastward speeds show a decrease with increasing latitude. The same is probably true of the westward speeds.

A significant variation of either eastward or westward speeds with season could not be detected. A much larger number of observations will be required, especially for the summer months, to establish seasonal characteristics.

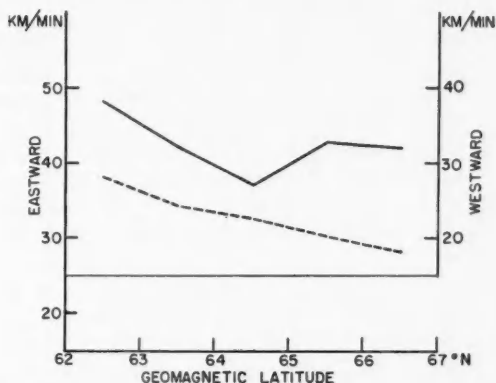


FIG. 2. Variation of average speeds, eastward and westward respectively, with geomagnetic latitude. The upper curve is for westward speeds; the lower, for eastward speeds.

Diurnal variations of direction and speed are indicated by Table III. Westward motions are characteristic of the early evening hours, eastward motions of the early morning hours. Eastward or westward motions may occur in the 4 to 5 hours centered on local midnight. Eastward speeds are a maximum between 03 and 04 hours, and westward speeds between 21 and 22 hours.

Figure 3 is a graph of the average speed of the west-east motions for 10 km/min intervals of speed against the corresponding average values of the

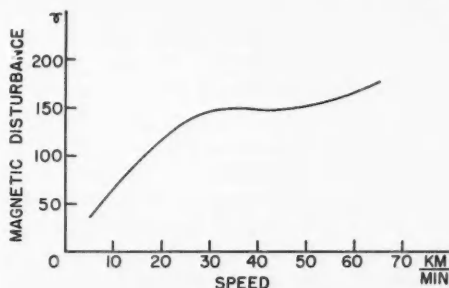


FIG. 3. Plot of the average departures from normal of the magnetic field intensity during observable, east-west motions of auroras for 10 km/min ranges of speed.

magnetic disturbance as measured by  $\Delta F$ . The speeds increase rapidly with moderate increases in the magnetic disturbance. Speeds above 30–40 km/min occur simultaneously with small or no further increases in magnetic disturbance above those observed for the lower speeds. Table III gives the diurnal variation in direction according to the number of occurrences for each hour of local time, and the corresponding numbers of positive and negative bays. A positive bay is one with an increase in the horizontal component of the earth's magnetic field; a negative bay, one with a decrease. Positive bays occur most frequently

TABLE III

Average speeds of the eastward and westward motions of aurora at Flin Flon grouped according to local mean time. The numbers in parentheses give the occurrences on which the averages are evaluated. The numbers of positive and negative magnetic bays are also given according to each hour of local time. For some observations the magnetic disturbance could not be designated as either positive or negative

L.M.T. (hours)	Eastward motion			Westward motion		
	Average speed (km/min)	Positive bays	Negative bays	Average speed (km/min)	Positive bays	Negative bays
18-19				23.1 (1)	1	
19-20				28.7 (5)	5	
21-22	38.4 (2)	1	1	33.3 (14)	12	
22-23	22.4 (8)	4	4	38.5 (18)	13	4
23-24	29.8 (8)	2	5	32.7 (29)	20	9
0-1	27.4 (25)	2	23	32.0 (39)	16	20
1-2	34.4 (20)	1	16	28.3 (34)	12	19
2-3	34.5 (21)	1	19	20.8 (9)	6	2
3-4	41.5 (9)		9			
4-5	35.3 (3)		3			
5-6	26.7 (2)		2			
6-7	25.2 (3)		3			

in the evening hours and with westward motions; negative bays, in the morning hours and with eastward motions. Positive or negative bays may occur with motions in either direction, but such cases are most frequent close to midnight when the direction of motion is changing.

#### DISCUSSION

The importance attached to the observed motions depends on several factors. The error in a horizontal displacement may be large because of the small size of the photograph (16 mm), the blurring of the edge of the auroral structure by the motion during the exposure (20 seconds), and distortion at low altitudes because of aberrations inherent in the camera. These should disappear from the averages based on a large number of observations, since a displacement larger than the correct one is as likely to be obtained as a smaller one. Motions parallel and normal to the geomagnetic meridians are almost certain to be simultaneously present in auroral structures. Homogeneous arcs and bands lack distinguishing features that persist for a sufficient time to detect the motion normal to the meridians. The transient nature of the convolutions and the rapidity of the east-west motion of the active forms suitable for measuring displacements normal to the magnetic meridians make it difficult to detect simultaneous displacements parallel to the meridians. In any case, the former is large compared to the latter. Displays are seldom orientated so that they are parallel to the geomagnetic lines of latitude. The angle between the geomagnetic meridians and the westward extension of homogeneous arcs (measured from south to west) averages about  $104^\circ$  according to Jensen and Currie (1953) for the region to the north of Saskatoon. As the magnetic disturbance increases this angle decreases, and displays extend east-west practically parallel to the lines of geomagnetic latitude. These

observations suggest that the most significant features of the horizontal motions is the north-south motion for quiet forms, and the east-west motion for active forms.

Since speeds are deduced on the assumption that the lower edge of an auroral structure is at a height of 100 km, the observed geographical, diurnal, and seasonal variations of speed may reflect nothing more than corresponding changes in height. Harang (1944) found that the average heights of the arcs over Norway decreased from about 114.3 to 94.7 km as their intensities increased from weak to very strong. Egedal (1937) has shown that heights decrease with time after sunset. Störmer (1955) found that ray-bands decrease in height to a minimum during the early morning hours, and that this height decrease is largest at distances of 24 to 26 degrees from the geomagnetic pole. Studies of the variation of auroral heights with magnetic disturbance do not appear to have been made, although it is generally accepted that the most active and intense auroras occur with large magnetic disturbance, suggesting a decrease of height at such times. The percentage error in a speed because of an error in the height can be computed for specific locations of the aurora relative to the photographic station. A deviation in height of  $\pm 20$  km from the assumed 100 km for aurora at an altitude of  $20^\circ$  above the horizon can lead to errors of about  $\pm 16\%$  in the deduced speeds; and for aurora at an altitude of  $90^\circ$ , the errors are about  $\pm 21\%$ . For actual heights larger than 100 km the computed speed is too small; for those smaller than 100 km, too large. For a large number of observations, the average speed should be substantially correct, if the average height does not depart appreciably from 100 km.

No obvious explanation exists for the observed distributions of speeds. Recent discoveries by Van Allen, McIlwain, and Ludwig (1958, 1959) and others of belts of trapped particles, spiralling forward and backward along the lines of the earth's magnetic field, suggest that most auroras (if not all) are caused by electrons scattered downward from these belts to auroral levels. These electrons, due to the trapping action of the earth's magnetic field, should participate in the rotation of the earth, and the resulting auroras should not show motions relative to a beam of charged particles entering the earth's atmosphere from space. According to an analysis of the motion by Welch and Whitaker (1959) the electrons should drift eastward relative to the earth's magnetic field, the rate of drift depending on the energies of the electrons. While only the most energetic electrons are likely to reach auroral levels, it is not apparent how the motions of these electrons relative to the magnetic field can result in the observed motions of auroras in either the north-south or the east-west directions.

The observed variations of speed with geomagnetic latitude and magnetic activity are probably real. The intensity of auroras, on an average, increase toward and in the auroral zone, and this should lead to a corresponding decrease in the average height relative to the assumed 100-km height. Hence the actual speeds should increase at a more rapid rate than the deduced speeds because of the failure to take into account the decreases in height. A similar

argument applies to the observed speeds and magnetic disturbance. If heights on an average decrease with increasing magnetic activity, the actual speeds should also increase at a more rapid rate than the deduced speeds. The existence of the diurnal variations of speed is less certain. A decrease in height with hours from sunset, and particularly in the morning hours, would result in a diurnal change of speed comparable with that deduced for Saskatoon and Flin Flon. The increase in the deduced speeds for the morning hours at Aklavik then imply an increase in auroral heights during the morning hours for stations inside the auroral zone. Height measurements by Currie (1955) at Chesterfield, N.W.T. ( $63.3^{\circ}$  N.,  $90.7^{\circ}$  W.) did not show such an increase, but many factors affect the height of auroras and large numbers of measurements are required to establish the magnitude of each one.

Horizontal motions in a north-south direction may be much greater for brief intervals of time than the ones observed with all-sky cameras. McEwen and Montalbetti (1958) report speeds up to 60 km/min from parallax measurements on auroras over Churchill ( $58.8^{\circ}$  N.,  $94.2^{\circ}$  W.). Their photographs were made at the rate of 10 per minute, making possible the detection of the fine structure in the speed spectrum. Speeds deduced from the all-sky camera photographs are effectively average values for 4-minute intervals, and never exceeded 48 km/min.

East-west motions of radar echoes from aurora in the Flin Flon region, measured by Lyon and Kavadas (1958), agree in broad detail with the motions of the photographic aurora. Their mean speeds of 6 and 15 km/min respectively for westward and eastward speeds are substantially less than the average speed of about 31 km/min for both directions found from the photographs. This is because only the most active forms have structures suitable for following the motion by visual or optical methods. Their observations also reduced the observed period in time during which motions in either direction can occur from 4-5 hours to 2 hours.

#### ACKNOWLEDGMENTS

The assistance of Mr. B. Morrison in operating a camera and magnetic variometers at Flin Flon, of Mr. D. Bekolay and Mr. D. Markowsky in operating a camera at Uranium City, and of Mr. F. Hector in operating a camera and magnetic variometers at Saskatoon is again gratefully acknowledged. The camera at Aklavik was placed there by the Geophysical Institute, College, Alaska, and operated by the Department of Northern Affairs. Financial assistance was provided by the Defence Research Board and the National Research Council of Canada.

#### REFERENCES

- CURRIE, B. W. 1955. *Can. J. Phys.* **33**, 773.  
EGEDAL, J. 1937. *As quoted by L. HARANG in The aurorae* (John Wiley & Sons, New York).  
FRITZ, H. 1881. *Das Polarlicht* (F. A. Brockhaus, Leipzig).  
HARANG, L. 1944. *Geophys. Publ.* **13**, No. 14.  
JENSEN, R. E. and CURRIE, B. W. 1953. *J. Geophys. Research*, **62**, 183.

- KIM, J. S. and CURRIE, B. W. 1958. *Can. J. Phys.* **36**, 160.  
LYON, G. F. and KAVADAS, A. 1958. *Can. J. Phys.* **36**, 1661.  
MC EWEN, D. J. and MONTALBETTI, R. 1958. *Can. J. Phys.* **36**, 1593.  
STÖRMER, C. 1955. *The polar aurora* (Clarendon Press, Oxford).  
VAN ALLEN, J. A., MCLWAIN, C. E., and LUDWIG, G. H. 1959. *J. Geophys. Research*, **64**, 271, 877.  
WEAVER, J. T. and SKINNER, R. 1960. *Can. J. Phys.* **38**, 1089.  
WELCH, J. A. and WHITAKER, W. A. 1959. *J. Geophys. Research*, **64**, 909.

# VISCOSITY MEASUREMENTS IN LIQUID HELIUM II<sup>1</sup>

C. B. BENSON<sup>2</sup> AND A. C. HOLLIS HALLETT

## ABSTRACT

Measurements of the viscosity of liquid helium II have been made using an oscillating sphere. This method avoids the necessity of a "corner" correction unavoidable when a circular disk is used, and therefore eliminates the uncertainty associated with such a correction. Calibration experiments showed the presence of a significant contribution to the observed damping of the oscillations which arose from the viscous drag of the gas surrounding the rod which connected the sphere with the torsion suspension fiber. This damping has been calculated and when applied to the results obtained in liquid helium II, the values of the viscosity of the normal component which were obtained agree with those obtained by the rotating cylinder method within the combined experimental uncertainties. The assumed density of the normal fluid was that obtained from the velocity of second sound, and the most accurate thermal data available.

## 1. INTRODUCTION

The viscosity,  $\eta_n$ , of the normal component of liquid helium II has been determined by a number of different methods. The oscillating disk method (Andronikashvili 1948; de Troyer *et al.* 1951; Hollis Hallett 1952; Dash and Taylor 1957), and the rotating cylinder viscometer (Heikkila and Hollis Hallett 1955; Woods and Hollis Hallett, to be published) are the direct methods. Indirect methods which have given values of the viscosity include measurements of the attenuation of second sound (Zinovleva 1957) and, more recently, some heat conduction experiments in capillary tubes (Brewer and Edwards 1959). With the exception of the oscillating disk measurements, all these experiments have given values of  $\eta_n$  which agree within the limits of the experimental errors. Of the oscillating disk measurements, the results of the earlier workers are in reasonable agreement among themselves, although not with the results of other methods; but the recent measurements by Dash and Taylor (1957) give values for the viscosity which are as much as 30% higher than the other measurements.

It has been suggested (Ginzburg 1956) that the discrepancy between the oscillating disk measurements and measurements by other methods may be due to the surface energy associated with the discontinuity which exists in the tangential velocity of the superfluid at a solid boundary. Another possible explanation is that since the oscillating disk method gives  $\eta\rho$ , the product of viscosity and density, and not  $\eta$  directly, the discrepancy may be accounted for if the values of the density,  $\rho_n$ , of the normal component, which are obtained from other experiments, are in error. Furthermore it was felt that the oscillating disk measurements were also subject to some uncertainty because the Navier-Stokes equation for the flow of a viscous fluid cannot be solved exactly for the oscillating disk; a correction for the damping due to the liquid situated

<sup>1</sup>Manuscript received July 4, 1960.

Contribution from the Department of Physics, University of Toronto, Toronto, Ont.

<sup>2</sup>Present address: Department of Physics, University of Ottawa, Ottawa 2, Ontario.



at the corners of the disk must be applied, and this is deduced either by Meyer's (1887) approximate treatment or by an empirical method (Dash and Taylor 1957). To avoid this last uncertainty it was decided to use an oscillating sphere rather than a disk, since no corner correction, empirical or otherwise, is necessary and the Navier-Stokes equation can therefore be solved more exactly.

During the course of the measurements, calibration experiments were carried out in low temperature fluids of known viscosity and density, and the results were somewhat inconsistent. It was discovered that the inconsistencies were due to the fact that no allowance had been made for the damping of the gas upon the rod which connected the sphere at the low temperature part of the cryostat to the torsion fiber, which was kept at room temperature. Accordingly the mathematical solution of the Navier-Stokes equation for the damping of the gas on the cylindrical rod was obtained, and the decrement due to this source calculated using reasonable assumptions as to the variation of temperature, and therefore of viscosity and density of the gas, along the length of the rod. When allowance for this damping due to the gas is made, the results obtained for the viscosity of helium II agree quite well with the rotating cylinder viscometer values. It is therefore felt that the previous discrepancies between oscillating disk and rotating cylinder values of the viscosity are, in most cases, due to a neglect of this damping due to the gas, and that there is no evidence to show a real discrepancy.

## 2. THE APPARATUS

Although the apparatus has already been described (Benson and Hollis Hallett 1956), a brief description will be given here emphasizing details particularly relevant to this work. It is illustrated schematically in Fig. 1. The sphere (A) was made of two spun copper hemispheres having a wall thickness of about 0.006 in. soldered together around the equator. The excess solder was removed, the surface smoothed, and the whole sphere was copper-plated to a thickness of 0.001 in. The diameter of the finished sphere did not vary by more than 1% from the mean which was 2.587 cm at room temperature, or 2.580 cm when corrected for contraction to helium temperatures (Fraser and Hollis Hallett 1955). A short piece of brass tubing was soldered to the north pole so that the straightened quartz rod (B) of length 85 cm and radius 0.11 cm could be attached to the sphere with a small amount of shellac. The upper end of the rod (B) was cemented with shellac to a small four-sided mirror which, in turn, was fastened to the lower end of the quartz suspension (C). The upper end of the suspension was secured to the top of the torsion head (D) and provision was there made for lateral centering of the point of suspension with respect to the axis of the Dewar. The whole of the quartz rod was enclosed by a thin-walled monel tube (E) of internal radius 0.53 cm which terminated in a plate (F) to which a chamber, completely enclosing the sphere, could be sealed. With this chamber in place, the space surrounding the sphere could be pumped out by means of a diffusion pump so that the "vacuum" decrement and period of the oscillating system could be determined when the sphere was at liquid helium temperatures.

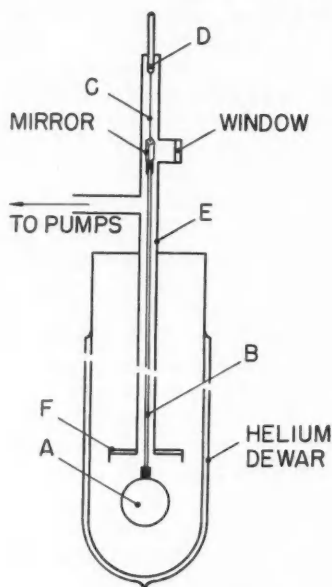


FIG. 1. The cryostat.

Observations of the torsional oscillations of the sphere were made using a beam of light reflected from the mirror and passing through a window fixed to the torsion head.

The equation used to calculate the viscosity,  $\eta$ , of the fluid surrounding the sphere from the observations of the logarithmic decrement,  $\Delta$ , and period,  $T$ , of the oscillations of the sphere is that given by Verschaffelt (1916).

$$(1) \quad \eta\rho = \frac{9I^2}{16\pi R^8} \frac{[\Delta(\tau^2+1)-2\tau\Delta_0]^2}{T(1-\Delta)} \left\{ 1 + \frac{2}{R\Lambda} + \frac{1+1/R\Lambda}{(R\Lambda)^2 + (1+R\Lambda)^2} \right\}^{-2}$$

where  $\tau = T/T_0$  and  $\Delta_0$  and  $T_0$  are respectively the logarithmic decrement and period when the sphere oscillates in a vacuum.  $R$  and  $I$  are respectively the radius and moment of inertia of the sphere and  $1/\Lambda = (2\eta/\omega\rho)^{1/2} = \lambda$  is the depth of penetration of the viscous waves propagated from the surface of the sphere into the fluid of viscosity  $\eta$  and density  $\rho$ . The moment of inertia of the system,  $I = 3.292 \text{ g cm}^2$  (at  $4^\circ \text{ K}$ ), was determined by comparison with disks of known moments of inertia, and the apparatus constant,  $9I^2/16\pi R^8 = K$ , was calculated to be  $0.253 \pm 0.005 \text{ c.g.s. units}$ ; this figure has been corrected for contraction of the sphere when cooled to liquid helium temperatures since the total contraction of copper is known (Fraser and Hollis Hallett 1955).

### 3. CALIBRATION OF THE SPHERE

During the course of these experiments it was decided to check the measurements of the apparatus constant  $K$  by measuring the observed decrement,

$\Delta_{\text{obs}}$ , and period of oscillation of the sphere when it was immersed in a fluid of known viscosity and density. Since the sphere was hollow and contained air it was necessary to use a fluid at a sufficiently low temperature that the air would solidify and not contribute any damping to the system. Helium gas at 4.22° K was selected for this purpose since the viscosity (van Itterbeek and Keesom 1938) and the density (Erickson and Roberts 1954) are known. The values of  $K$  determined by these calibration experiments by substituting  $\Delta_{\text{obs}}$  for  $\Delta$  in equation (1) at three different periods of oscillation are shown in Table I. When these values are compared with the calculated value of  $K = 0.253 \pm .005$  c.g.s. units, it can be readily seen that not only does the

TABLE I  
The apparatus constant,  $K$ , found by calibration of sphere in  
helium gas at 4.22° K

Period $T$ (sec)	$K$	$10^3 \Delta_{\text{obs}}$	$10^3 \Delta_0$	$10^3 \Delta_g$
6.7	0.195	1.48	0.06	0.19
17.6	0.168	3.02	0.44	0.50
24.5	0.151	3.94	0.70	0.76

calibration experiment give appreciably lower values of  $K$ , but also gives values which decrease as the period of oscillation was increased. Similar results were also found using helium I and liquid hydrogen as calibrating fluids.

These discordant values of  $K$  would seem to point to one of two possibilities: either equation (1) is wholly inadequate or some quantity appearing in equation (1) has been incorrectly measured. The first possibility seems most unlikely since the equation has been derived by many others (see, for example, Kestin and Wang 1955; Mariens and Scheltgens 1957) and the result obtained by each differs insignificantly from Verschaffelt's result given in equation (1). Consideration of the experimental situation does, however, lead to a possibility that the gas surrounding the rod (B) of Fig. 1 might contribute a significant amount to the total observed damping and that therefore the quantity  $\Delta$  has been incorrectly measured. Denoting by  $\Delta_{\text{obs}}$  and  $\Delta_g$  the observed decrement and the decrement due to the damping of the gas on the supporting rod (B) respectively, one can write  $\Delta_{\text{obs}} = \Delta + \Delta_g + \Delta_0$  where  $\Delta$  is the logarithmic decrement due to the viscous forces of the fluid on the sphere itself and is the quantity required by equation (1). If the discordant results of the calibration experiments are now interpreted as the neglect of the gas decrement it is possible to calculate  $\Delta_g$  from these results assuming that the apparatus constant  $K$  has the value 0.253 c.g.s. units determined by measurements of its mechanical properties. The results are included in Table I together with values of  $\Delta_{\text{obs}}$  and  $\Delta_0$ . It can be seen that  $\Delta_g$  can become of the order of 20% of the observed decrement.

The question now arises as to whether such a large gas decrement is physically reasonable and whether its value can be estimated reliably and applied to results obtained in liquid helium II. Accordingly the Navier-Stokes equation

has been solved for the case of a cylindrical rod (the supporting rod (B) of Fig. 1) oscillating within a fixed cylindrical wall (the tube (E) of Fig. 1) and the details of this calculation are given in the Appendix. An immediate problem arises in the application of this solution of the Navier-Stokes equation to the calculation of the damping couple and thus the gas decrement  $\Delta_g$  for the present experimental situation in which the temperature of the gas, and therefore its viscosity and density, varies from helium temperatures at the bottom of the rod to room temperature at the top. It is well known that most of the variation of temperature takes place near the room temperature part of the apparatus, and for a reasonable approximation it has been assumed that the temperature of the gas is 300° K for the first 15 cm down from the join with the suspension fiber, 90° K for the next 15 cm, and 4° K for the next 50 cm. The remaining 5 cm of the 85-cm length of the rod is usually immersed in the liquid helium and does not appear to contribute significant damping since no variation of the observed decrement could be detected as the bath level fell, thereby decreasing the length of the rod immersed.

The result of this analysis shows clearly that values of  $\Delta_g$  such as those given in Table I are perfectly possible and arise largely from the gas damping at the room temperature end of the rod. Such an effect as gas damping could therefore, if neglected, lead to spurious values of the viscosity deduced from measurements of the decrement and period of torsion pendulums. Accordingly equation (1) should be rewritten as

$$(2) \quad \eta\rho = \frac{9I^2}{16\pi R^5} \frac{[\Delta_{\text{obs}}(\tau^2+1) - 2\tau\Delta_g]^2}{T(1-\Delta_{\text{obs}})} \left\{ 1 + \frac{2}{RA} + \frac{1+1/RA}{(RA)^2 + (1+RA)^2} \right\}^{-2}$$

where  $\Delta_{\text{obs}}$  is the observed decrement, and  $\Delta_g$  has been written for  $\Delta_0 + \Delta_g$ . The numerical calculations have been carried out for the various pressures of helium gas in the Dewar corresponding to various temperatures of the helium II bath, and at the various periods of oscillation used in this work. Applying these results to measurements made of the decrement and period of the sphere while oscillating in helium II values of the viscosity are calculated which are given in Table II together with the other relevant data. The values of  $\Delta_g$  taken from the Appendix are included in each section of Table II. The errors given in the final column have been calculated assuming that  $\Delta_g$  is the only source of error and has an uncertainty of 33½%. It is difficult to estimate exactly what the temperature distribution along the rod actually is during an experiment, but it is felt that with all effects considered 33½% uncertainty in the calculation of  $\Delta_g$  is certainly not unreasonable.

#### 4. DISCUSSION OF RESULTS

The viscosity values given in Table II are plotted against temperature in Fig. 2 and are there compared with the results obtained using the rotating cylinder viscometer (Woods and Hollis Hallett, to be published). The present results lie much closer to the rotating cylinder results than any of the previous oscillating disk results (Andronikashvili 1948; de Troyer *et al.* 1951; Hollis Hallett 1952). In fact, it is felt that with the uncertainties in the estimation

TABLE II

Experimental data for sphere with calculated viscosity. Density,  $\rho_n$ , found by interpolation from Table III

$T$ ( $^{\circ}\text{K}$ )	Period $T$ (sec)	$10^3\Delta$	$\rho_n$ (g/cc)	$10^6\eta_n$ (poise)
(a)				
2.169	6.472	5.099	.1419	$26.3 \pm 0.4$
2.088	6.461	3.790	.1079	$18.6 \pm 0.4$
1.993	6.459	3.089	.0811	$16.1 \pm 0.4$
1.821	6.454	2.320	.0490	$14.3 \pm 0.5$
1.651	6.452	1.861	.0286	$14.9 \pm 0.7$
1.489	6.449	1.503	.01595	$16.3 \pm 0.9$
1.357	6.447	1.235	.00908	$17.7 \pm 1.3$
1.214	6.447	0.976	.00440	$19.8 \pm 1.8$
1.083	6.445	0.842	.00195	$27.7 \pm 3.0$
$T_0 = 6.440 \text{ sec } \Delta_s = \Delta_0 + \Delta_g = (0.056 + 0.118) \times 10^{-3}$				
(b)				
2.170	6.675	5.202	.1427	$26.4 \pm 0.4$
2.091	6.666	3.879	.1090	$18.7 \pm 0.4$
2.003	6.661	3.109	.0835	$15.3 \pm 0.4$
1.816	6.656	2.344	.0483	$14.3 \pm 0.5$
1.648	6.653	1.889	.0283	$15.0 \pm 0.7$
1.505	6.652	1.565	.0170	$16.1 \pm 0.8$
1.402	6.651	1.357	.0111	$17.5 \pm 1.1$
1.309	6.649	1.210	.00723	$20.0 \pm 1.4$
1.211	6.648	1.025	.00433	$21.6 \pm 1.8$
1.102	6.647	0.879	.00222	$26.4 \pm 2.7$
$T_0 = 6.640 \text{ sec } \Delta_s = \Delta_0 + \Delta_g = (0.056 + 0.122) \times 10^{-3}$				
(c)				
2.168	23.07	9.995	.1411	$26.1 \pm 0.7$
2.161	23.06	9.805	.1363	$25.8 \pm 0.7$
1.991	22.98	6.263	.0807	$16.1 \pm 0.7$
1.812	22.95	4.806	.0477	$14.5 \pm 0.9$
1.652	22.93	3.998	.0287	$15.1 \pm 1.2$
1.412	22.89	3.028	.0115	$17.4 \pm 1.9$
1.262	22.86	2.563	.00569	$20.5 \pm 2.7$
1.121	22.84	2.143	.00252	$23.6 \pm 3.8$
$T_0 = 22.80 \text{ sec } \Delta_s = (\Delta_0 + \Delta_g) = (0.310 + 0.422) \times 10^{-3}$				
(d)				
2.137	23.10	8.863	.1255	$22.5 \pm 0.7$
2.095	23.08	7.783	.1104	$19.2 \pm 0.7$
1.917	23.03	5.545	.0651	$15.0 \pm 0.8$
1.595	22.99	3.746	.0236	$15.4 \pm 1.3$
1.469	22.98	3.242	.0147	$16.6 \pm 1.6$
1.319	22.96	2.717	.00758	$18.8 \pm 2.3$
1.108	22.95	2.102	.00231	$23.7 \pm 3.9$
$T_0 = 22.87 \text{ sec } \Delta_s = (\Delta_0 + \Delta_g) = (0.310 + 0.423) \times 10^{-3}$				

NOTE: Radius of sphere = 1.290 cm corrected for contraction.

Moment of inertia of system = 3.292 c.g.s. units, corrected for contraction.

of the gas damping, there is no reason to suppose that there is any real discrepancy between the oscillating sphere and the rotating cylinder results. None of these previous oscillating disk results made any allowance for gas damping and it is suggested that the discrepancies between the rotating cylinder and the oscillating disk values of the viscosity may, in large measure, be due to neglect of the gas damping both in the actual measurements and in any calibration experiment.

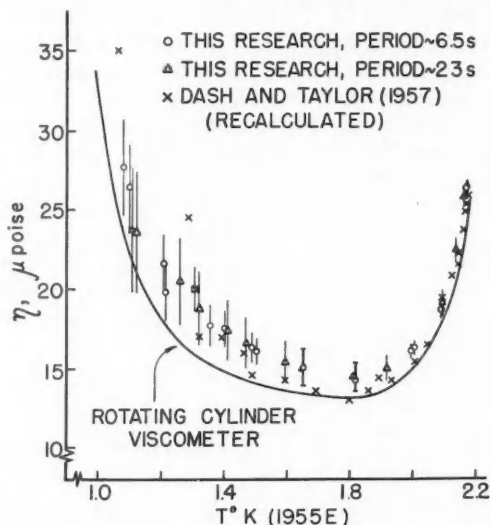


FIG. 2. The viscosity,  $\eta_n$ , as a function of temperature.  $\circ$ , oscillating sphere ( $T \sim 6$  sec);  $\Delta$ , oscillating sphere ( $T \sim 22$  sec);  $\times$ , oscillating disk (Dash and Taylor 1957) recalculated using  $\rho_n$  as given in Table III; —, rotating cylinder (Heikkila and Hollis Hallett 1955; Woods and Hollis Hallett, to be published).

A further problem arises with the oscillating disk, namely, that the appropriate solution of the Navier-Stokes equation is less exact for the disk than it is for the sphere since approximations are necessary to take into account the damping due to the liquid situated at the sharp corners of the disk. An empirical correction for this corner damping was given by Dash and Taylor (1957) which appears more satisfactory than any of the previous corner corrections, but it should be pointed out that the sphere required no such correction, empirical or otherwise, and is a fundamentally more satisfactory geometrical figure.

The values of  $\rho_n$ , the density of the normal fluid in liquid helium II, which are given in Table II have been taken from the data given in Table III. These values have been calculated using the Landau equation which connects the ratio  $\rho_n/\rho$  with the velocity of second sound  $u_2$  and specific heat  $C$  and entropy  $S$  of liquid helium II. Values of  $u_2$  were obtained from a large scale plot of  $u_2$  versus temperature including all available published values, and values of  $C$  and  $S$  were taken from the accepted smoothed values which accompany the vapor pressure table of the 1955L scale (van Dijk and Durieux 1956). The resulting calculated values of  $\rho_n/\rho$  were smoothed by eye until the first and second differences were reasonably smooth functions of temperature. Finally  $\rho_n$  was calculated using for  $\rho$  the values given by Kerr (1957). The maximum adjustment of the calculated values of which was made to produce the smoothed values in Table III was 2.7%.

TABLE III  
The density of the normal component in helium II

$T$ ( $^{\circ}\text{K}$ )	$\rho_n$ ( $\text{g cm}^{-3}$ )	$T$ ( $^{\circ}\text{K}$ )	$\rho_n$ ( $\text{g cm}^{-3}$ )
2.173 ( $T_\lambda$ )	0.1466	1.75	0.0394
2.170	0.1427	1.70	0.0336
2.160	0.1357	1.65	0.0285
2.150	0.1308	1.60	0.0240
2.140	0.1267	1.55	0.0201
2.130	0.1229	1.50	0.01665
2.120	0.1192	1.45	0.0136
2.110	0.1156	1.40	0.0110
2.100	0.1121	1.35	0.0088
2.05	0.0960	1.30	0.0069
2.00	0.0828	1.25	0.00535
1.95	0.0716	1.20	0.00407
1.90	0.0620	1.15	0.00303
1.85	0.0535	1.10	0.00219
1.80	0.0460	1.05	0.00153
		1.00	0.00103

$\rho_n$  calculated from  $u_2^2 = (\rho - \rho_n)TS^2/\rho_n C$  (Landau 1941),

$C$  and  $S$  from van Dijk and Durieux (1956),

$\rho$  from Kerr (1957),

$u_2$  from smoothed values using all published data.

Temperatures are on the  $T_{15}\text{K}$  scale (Clement *et al.* 1955).

The recent results of Dash and Taylor (1957) require special mention. They employed the oscillating disk and considered the possibility of gas damping but stated that it was negligible. Calculations of  $\Delta_g$  making similar assumptions with regard to temperature distribution bear out this statement and show that for their apparatus  $\Delta_g$  would be very much smaller than in the present experiments, chiefly because the radius of the quartz rod connecting the disk with the torsion fiber is smaller, and would only amount to about 0.3% of the total decrement at the lowest temperatures. They also determined empirically a correction for the damping due to the liquid at the corners of the disk. Their final results are very self-consistent, but the viscosity values are all larger than the rotating cylinder values. If, however, their results are recalculated using the data for  $\rho_n$  which are given in Table III, it is found that the values of the viscosity so calculated are in good agreement with the rotating cylinder and the present oscillating sphere results, as shown also in Fig. 2. Their values of  $\rho_n$  were determined by the Andronikashvili (1948) method but are smaller than the values given in Table III and also than those obtained by calculations using the spectrum of the excitations of liquid helium II (Bendt *et al.* 1959). It is suggested that the values of  $\rho_n$  determined by Dash and Taylor are, for some reason, in error and that the data given in Table III are closer to the real values.

It is therefore concluded that the damping of the gas on the rod supporting the bob of the torsion pendulum can contribute appreciably to the measured logarithmic decrement and, if neglected, can lead to values of the viscosity which are too large. Within the errors introduced by the estimation of this gas damping, the viscosity values obtained using the oscillating sphere agree with the results obtained using the rotating cylinder viscometer.

## ACKNOWLEDGMENTS

The authors wish to acknowledge the grants-in-aid of research provided by the National Research Council of Canada and the Ontario Research Foundation without which this work would not have been possible. The authors also gratefully acknowledge the generous assistance of the Computation Centers of the University of Toronto and of the University of Ottawa in the numerical calculations.

## REFERENCES

- ANDRONIKASHVILI, E. L. 1948. *J. Exptl. Theoret. Phys. (U.S.S.R.)* **18**, 424, 429.  
 BENDT, P. J., COWAN, R. D., and YARNELL, J. L. 1959. *Phys. Rev.* **113**, 1386.  
 BENSON, C. B. and HOLLIS HALLETT, A. C. 1956. *Can. J. Phys.* **34**, 668.  
 BREWER, D. F. and EDWARDS, D. O. 1959. *Proc. Roy. Soc. (London)*, A, **251**, 247.  
 CLEMENT, J. R., LOGAN, J. K., and GAFFNEY, J. 1955. *Phys. Rev.* **100**, 743. See their "Note added in proof".  
 DASH, J. G. and TAYLOR, R. D. 1957. *Phys. Rev.* **105**, 7.  
 DE TROYER, A., VAN IITERBEEK, A., and VAN DEN BERG, G. J. 1951. *Physica*, **17**, 50.  
 ERICKSON, R. A. and ROBERTS, L. D. 1954. *Phys. Rev.* **93**, 957.  
 FRASER, D. and HOLLIS HALLETT, A. C. 1955. *Bull. inst. intern. froid*, **35**, 790.  
 GINZBURG, V. L. 1956. *Soviet Physics, JETP*, **2**, 170.  
 HEIKKILA, W. J. and HOLLIS HALLETT, A. C. 1955. *Can. J. Phys.* **33**, 420.  
 HOLLIS HALLETT, A. C. 1952. *Proc. Roy. Soc. (London)*, A, **210**, 404.  
 KESOM, W. H. 1942. *Helium* (Elsevier, Amsterdam), Chap. 2, p. 106.  
 KERR, E. C. 1957. *J. Chem. Phys.* **26**, 511.  
 KESTIN, J. and WANG, H.-E. 1955. *Brown Univ. Report AF-891/5*.  
 LANDAU, L. D. 1941. *J. Phys. U.S.S.R. (Moscow)*, **5**, 71.  
 MARIENS, P. and SCHELTEGERS, R. 1957. *Mededeel. Koninkl. Vlaam. Acad. Wetenschap. Belg.*, XIX, 3.  
 MEYER, O. E. 1861. *J. reine angew. Math.* **59**, 229.  
 ——— 1887. *Ann. Physik (Wied)*, **32**, 642.  
 VAN DIJK, H. and DURIEUX, M. 1956. *Physica*, **22**, 760.  
 VAN IITERBEEK, A. and KESOM, W. H. 1938. *Physica*, **5**, 257.  
 VERSCHAFFELT, J. E. 1916. *Leiden Comm.* 148b.  
 WOODS, A. D. B. and HOLLIS HALLETT, A. C. To be published.  
 ZINOVEVA, K. N. 1957. *Soviet Physics, JETP*, **4**, 36.

## APPENDIX

Consider a rod of circular section, radius  $R$ , coaxial with a cylindrical tube of inside radius  $R'$ ; the space between the rod and the wall of the tube being filled with fluid of viscosity  $\eta$  and density  $\rho$ . Then it may be shown (Meyer 1861) that the angular velocity  $\psi_r$  of the fluid at a radius  $r$  intermediate between  $R$  and  $R'$  is given by the equation

$$(1) \quad \frac{\partial^2 \psi_r}{\partial r^2} + \frac{3}{r} \frac{\partial \psi_r}{\partial r} = \frac{\rho}{\eta} \frac{\partial \psi_r}{\partial t}$$

where it is assumed that the fluid is incompressible and the pressure gradient negligible so that  $\psi_r$  is a function of  $r$  and  $t$  only.

Consider further that the rod oscillates about its axis with a damped simple harmonic motion according to the equation

$$(2) \quad \alpha = \phi_0 e^{\theta t}$$

where  $\alpha$  is the angular displacement of the rod at time  $t$ ,  $\phi_0$  the initial amplitude, and  $\theta = \omega(i - \Delta)$ , where  $\omega = 2\pi/T$  if  $T$  is the period of oscillation, and  $\Delta$  is the logarithmic decrement.  $\alpha$  must satisfy the equation.



$$(3) \quad I\ddot{\alpha} + L\dot{\alpha} + K\alpha = 0$$

where  $I$  is the moment of inertia of the rod and attached system and  $K$  is the torsion constant of the fiber by which the system is suspended.  $L\dot{\alpha}$  is the total damping couple exerted on the system and is determined by the total viscous drag of the fluid on the rod by the equation

$$(4) \quad L\dot{\alpha} = \int 2\pi R^3 \eta (\partial \psi_r / \partial r)_{r=R} dl$$

where the integral is taken over the length of the rod.

It is required to solve equation (1) subject to the boundary conditions of the problem and thereby obtain  $(\partial \psi_r / \partial r)_{r=R}$ ;  $L$  can then be determined and the related logarithmic decrement deduced.

The solution of equation (1) is

$$(5) \quad \psi_r = \frac{1}{r} e^{\beta^2 \eta t / \rho} Z_1(i\beta r)$$

where  $\beta^2$  is a constant and  $Z_1(i\beta r)$  is given by

$$(6) \quad Z_1(i\beta r) = C_1 J_1(i\beta r) + C_2 N_1(i\beta r),$$

$J_1$  and  $N_1$  being Bessel functions of the first and second kind respectively, while  $C_1$  and  $C_2$  are constants to be determined from the boundary conditions. The boundary conditions are  $\psi_r \equiv 0$  for all  $t$  at  $r = R'$ , and  $\psi_r = \dot{\alpha} = \theta \phi_0 e^{\theta t}$  at  $r = R$ , assuming no slip at the surfaces.

Applying these boundary conditions to equations (5) and (6) it is found that

$$(7) \quad C_2 = -C_1 \frac{J_1(i\beta R')}{N_1(i\beta R')}$$

and

$$(8) \quad C_1 = R\phi_0 \theta \left\{ J_1(i\beta R) - \frac{J_1(i\beta R')}{N_1(i\beta R')} N_1(i\beta R) \right\}^{-1}$$

where

$$(9) \quad \theta = \omega(i - \Delta) = \beta^2(\eta/\rho).$$

Thus

$$(10) \quad \psi_r(r, t) = \frac{R\phi_0 \theta e^{\theta t}}{r} \left\{ \frac{J_1(i\beta r) N_1(i\beta R') - J_1(i\beta R') N_1(i\beta r)}{J_1(i\beta R) N_1(i\beta R') - J_1(i\beta R') N_1(i\beta R)} \right\}$$

and

$$(11a) \quad \left( \frac{\partial \psi_r}{\partial r} \right)_{r=R} = -i\beta \phi_0 \theta e^{\theta t} \left\{ \frac{J_2(i\beta R) N_1(i\beta R') - N_2(i\beta R) J_1(i\beta R')}{J_1(i\beta R) N_1(i\beta R') - J_1(i\beta R') N_1(i\beta R)} \right\}$$

which, for convenience, can be written

$$(11b) \quad \left( \frac{\partial \psi_r}{\partial r} \right)_{r=R} = -i\beta \phi_0 \theta e^{\theta t} (x + iy).$$

Furthermore, from equation (9) it follows that

$$(12) \quad \beta = \Lambda \{ (1 - \Delta)^{\frac{1}{2}} + i(1 + \Delta)^{\frac{1}{2}} \}$$

where

$$(13) \quad \Lambda = (\omega \rho / 2\eta)^{\frac{1}{2}}$$

and  $\Delta^2$  (usually  $\sim 10^{-3}$  in these experiments) has been neglected compared with unity.

The logarithmic decrement  $\Delta_g$  resulting from the damping couple exerted on the rod by the vapor follows from equations (4), (3), and (2) giving the result

$$(14) \quad 2I\omega\Delta_g = \mathcal{R}(L)$$

where  $\mathcal{R}(L)$  is written for the real part of the coefficient  $L$  given by equation (4).

It is convenient to consider as total nuisance decrement  $\Delta_s = \Delta_0 + \Delta_g$ , thereby combining the vacuum decrement,  $\Delta_0$ , due to internal friction in the torsion fiber, and the vapor decrement,  $\Delta_g$ , into a single quantity. It follows that

$$(15) \quad 2I\omega\Delta_s = 2I\omega_0\Delta_0 + 2I\omega\Delta_g$$

where  $\omega_0 = 2\pi/T_0$  if  $T_0$  is the period of oscillation in vacuum.

$$\text{From (15)} \quad \Delta_s = \Delta_0 \left( \frac{\omega_0}{\omega} \right) + \frac{1}{2I\omega} \mathcal{R}(L)$$

and from equations (4) and (11)

$$(16) \quad \Delta_s = \Delta_0 \left( \frac{\omega_0}{\omega} \right) + \frac{1}{2I\omega\alpha} \int 2\pi R^3 \eta [-i\beta \phi_0 e^{\theta t} (x + iy)] dl$$

where the real part of the integral is taken over the length of the rod. Substituting for  $\beta$  from equation (12), and making the assumptions that  $\omega \approx \omega_0$  and that  $\Delta \ll 1$  so that  $\beta = \Lambda(1 + i)$ , equation (16) becomes

$$(17) \quad \Delta_s = \Delta_0 - \frac{\pi R^2}{I\omega} \int \eta \Lambda R (x + y) dl.$$

Since both  $x$  and  $y$  are negative a positive contribution to  $\Delta_s$  will result.

To evaluate  $\Delta_g$  from equation (17) the quantities  $x$ ,  $y$ , and  $\Lambda$  must be determined. From equation (13) it follows that  $\Lambda$  depends upon the density  $\rho$  and viscosity  $\eta$  of the vapor surrounding the supporting rod, and therefore upon the distribution of temperature along this rod. This distribution is not known at all exactly, but it is known that for a cryostat having the geometrical arrangement shown in Fig. 1 the temperature decreases from room temperature (300° K) to a value near 4° K in a relatively short distance measured down from the top of the cryostat. It therefore seems reasonable to divide the length of the rod into three parts,  $l_1$ ,  $l_2$ , and  $l_3$  and to assume that

- $l_1$  the upper 15 cm of the rod is at 300° K,  $\eta_1 = 190 \times 10^{-6}$  poise;
- $l_2$  the next 15 cm of the rod is at 90° K,  $\eta_2 = 87 \times 10^{-6}$  poise;
- $l_3$  the lower 50 cm of the rod is at 4° K,  $\eta_3 = 13 \times 10^{-6}$  poise.

Equation (17) may then be rewritten as a sum rather than an integral.

$$(18) \quad \Delta_s = \Delta_0 - \frac{\pi R^3}{I\omega} [\eta_1 \Lambda_1 (x_1 + y_1) l_1 + \eta_2 \Lambda_2 (x_2 + y_2) l_2 + \eta_3 \Lambda_3 (x_3 + y_3) l_3]$$

where the subscripts indicate that the values of  $\eta$ ,  $\Lambda$ ,  $x$ , and  $y$  are taken at the temperature corresponding to the regions defined by  $l_1$ ,  $l_2$ , and  $l_3$ . The values of  $\eta_1$ ,  $\eta_2$ , and  $\eta_3$  given above are the viscosities of helium gas at the temperatures quoted. They are independent of the density of the gas, and are taken from Keesom (1942).

The density of the gas is calculated from the saturated vapor density measurements of Erickson and Roberts (1954) on the assumption that the pressure is constant along the length of the rod and equal to the vapor pressure of the helium bath, and also that at the low pressures involved the density will vary inversely with the temperature. Because the density,  $\rho$ , is determined by the vapor pressure of the helium bath it is convenient to quote values of  $\rho_1$ ,  $\rho_2$ , and  $\rho_3$  for the three temperature regions of the rod at various working temperatures of the helium bath. These values are given in Table IA.

TABLE IA

Values of the density  $\rho$  of the helium vapor and  $\Lambda = (\omega\rho/2\eta)^{1/2}$  at various temperatures in the three regions,  $l_1$ ,  $l_2$ , and  $l_3$  of the supporting rod

	Temperature of the helium bath, °K				
	2.15	2.0	1.8	1.5	1.2
$10^6 \rho_1$ (g cm <sup>-3</sup> )	0.83	0.53	0.28	0.075	0.0133
$\Lambda_1^6$ (cm <sup>-1</sup> )	0.145	0.116	0.084	0.0435	0.0183
$\Lambda_1^{22}$ (cm <sup>-1</sup> )	0.0775	0.0619	0.0469	0.0233	0.00983
$10^6 \rho_2$ (g cm <sup>-3</sup> )	2.8	1.8	0.92	0.25	0.044
$\Lambda_2^6$ (cm <sup>-1</sup> )	0.394	0.314	0.226	0.117	0.0496
$\Lambda_2^{22}$ (cm <sup>-1</sup> )	0.210	0.169	0.120	0.0627	0.0265
$10^6 \rho_3$ (g cm <sup>-3</sup> )	62	40	21	5.6	1.0
$\Lambda_3^6$ (cm <sup>-1</sup> )	4.80	3.84	2.78	1.44	0.610
$\Lambda_3^{22}$ (cm <sup>-1</sup> )	2.56	2.06	1.49	0.770	0.324

Also given in Table IA are values of  $\Lambda$  calculated according to equation (13) for the two periods of oscillation, 6.54 seconds and 22.8 seconds, used in these experiments. The subscript following the symbol  $\Lambda$  indicates the region,  $l_1$ ,  $l_2$ , or  $l_3$  for which the value of  $\Lambda$  was calculated and the superscripts 6 or 22 indicate that the value was calculated for a period of 6.54 seconds or 22.8 seconds.

The quantities  $x$  and  $y$  defined by equations (11a) and (11b) were calculated using the I.B.M. 650 digital computer at the University of Ottawa and the results are given in Table IIA. The values of  $R$  and  $R'$  used were those applicable to these experiments, namely 0.11 cm and 0.53 cm respectively. As in the case of  $\Lambda$  the subscripts applied to  $x$  and  $y$  indicate the region, and the superscripts the period of oscillation, for which the calculations were made.

TABLE IIA

$$(x+iy) = \frac{\{J_2(i\beta R)N_1(i\beta R') - N_2(i\beta R)J_1(i\beta R')\}}{\{J_1(i\beta R)N_1(i\beta R') - N_1(i\beta R)J_1(i\beta R')\}}$$

calculated using  $\beta = \Lambda(1+i)$  for the values of  $\Lambda$  given in Table IA.  $R = 0.11$  cm;  $R' = 0.53$  cm

	Temperature of the helium bath, °K				
	2.15	2.0	1.8	1.5	1.2
$x_1^6$	-65.69	-82.24	-113.0	-218.2	-519.8
$y_1^6$	-65.72	-82.26	-113.0	-218.2	-519.8
$x_2^6$	-24.09	-30.26	-42.10	-81.01	-191.7
$y_2^6$	-24.17	-30.33	-42.15	-81.04	-191.7
$x_3^6$	-1.708	-2.199	-3.172	-6.469	-15.51
$y_3^6$	-2.519	-2.922	-3.732	-6.767	-15.64
$x_1^{22}$	-122.5	-153.5	-202.6	-408.3	-967.5
$y_1^{22}$	-122.5	-153.5	-202.6	-408.3	-967.5
$x_2^{22}$	-45.23	-56.16	-78.90	-151.5	-359.2
$y_2^{22}$	-45.27	-56.21	-78.93	-151.5	-359.2
$x_3^{22}$	-3.474	-4.427	-6.224	-12.28	-29.31
$y_3^{22}$	-3.995	-4.850	-6.533	-12.43	-29.38

Substitution of these values of  $x$  and  $y$ , and of  $\Lambda$ ,  $\eta$ , and  $l$  into equation (18) gives values of  $\Delta_s$  and of  $\Delta_g = \Delta_s - \Delta_0$ , and these values are given in Table IIIA for the two periods of oscillation at various bath temperatures.

There are two interesting features of these results which should be noted. First, the largest contribution to  $\Delta_g$  arises from the room temperature region  $l_1$  because  $\eta$ ,  $x$ , and  $y$  all have the largest values in this region.  $\Lambda$  alone is

TABLE IIIA

$\Delta_g$  and  $\Delta_s = \Delta_0 + \Delta_g$  at various temperatures

Bath temperature, °K	$10^4 \Delta_g$	$10^4 \Delta_s$
For $T = 6.54$ sec $\Delta_0 = 0.56 \times 10^{-4}$		
2.15	1.22	1.78
2.0	1.21	1.77
1.8	1.21	1.77
1.5	1.205	1.765
1.2	1.205	1.765
For $T = 22.8$ sec $\Delta_0 = 3.10 \times 10^{-4}$		
2.15	4.22	7.32
2.0	4.22	7.32
1.8	4.22	7.32
1.5	4.22	7.32
1.2	4.22	7.32

smaller in the region  $l_1$  than in any other, but the reduction of  $\Lambda$  is more than compensated for by the increase in  $x$  and  $y$ . Thus the major error in the calculation of  $\Delta_g$  lies in the estimation of the length of the supporting rod, which is at room temperature. It should be further noted that from

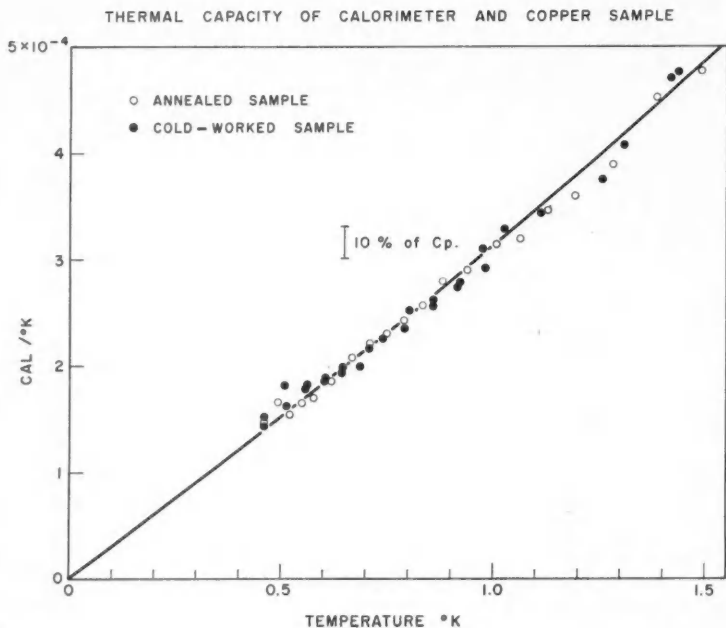
equation (18)  $\Delta_g$  is proportional to  $R^3$  where  $R$  is the radius of the rod. Thus a substantial reduction in  $\Delta_g$  can be achieved from a moderate reduction in the diameter of the supporting rod. The second interesting feature is that  $\Delta_g$  is virtually independent of the pressure in the cryostat and therefore of the helium bath temperature. This arises from the fact that  $-(x+y)$  appears to be inversely proportional to  $\Lambda$ . This has a practical significance, namely, that  $\Delta_g$  will depend only on the geometry of the cryostat and the period of oscillation, and will be relatively independent of the temperature of the bath and the pressure of the gas.

## NOTES

### THE SPECIFIC HEAT OF ANNEALED AND COLD-WORKED COPPER FROM 0.4° TO 1.5° K\*

DOUGLAS L. MARTIN

Previous measurements in the temperature range 20° K to 300° K (Martin 1960) showed that the specific heat of a heavily cold-worked sample of 99.999% pure copper was on the average about 0.15% higher than an annealed sample of the same material. At lower temperatures it was thought that differences in the lattice specific heat might become more marked since the predominating long wavelength phonons would be dependent upon the macroscopic rather than the microscopic elastic constants of the material. There is probably also a contribution to the specific heat due to the vibration of mobile dislocations (Granato 1958) which would vary linearly with temperature in the region of this experiment. However, with decreasing temperature the electronic com-



\*Issued as N.R.C. No. 5908.

Can. J. Phys. Vol. 38 (1960)

ponent of the specific heat becomes greater than the lattice component and the precision of measurement of the over-all specific heat decreases. Thus differences in the lattice component of the specific heat would have to be rather large in order to be observable with the present technique.

The copper samples were the same as had been used in the work at higher temperatures (Martin 1960). The mass of the annealed sample had been reduced to 60.470 g by etching and the cold-worked sample was turned to almost exactly the same mass (60.464 g). The cylindrical samples were clamped to a small copper block carrying the carbon thermometer (Speer) and heater by means of a band of 0.005 in. thick copper sheet. The measurements were made in a  $^3\text{He}$  cryostat using a superconducting heat switch for cooling. The temperature scale was based on the Los Alamos  $^3\text{He}$  vapor pressure data (Sydoriak and Roberts 1957).

The results are shown in Fig. 1. About 60% of the total thermal capacity is due to the copper sample. It seems clear that any difference in the thermal capacity of the two samples must be less than a few per cent. The lattice component of the specific heat varies from about 15% of the total at 1.5° K to only 2% at 0.5° K and thus quite large changes would be possible within the limits assigned to the possible variation in the total specific heat.

If the postulated term due to the oscillation of mobile dislocations does occur then its magnitude cannot exceed a few per cent of the electronic specific heat. Using the method proposed by Granato (1958) it may be inferred that the dislocation density is not greater than  $10^{11}$  dislocations per  $\text{cm}^2$ , in agreement with the observations of Blewitt, Coltman, and Redman (1954) on cold-worked copper.

- BLEWITT, T. H., COLTMAN, R. R., and REDMAN, J. K. 1954. Report of the Bristol Conference on Defects in Crystalline Solids. Physical Society, London, p. 369.  
GRANATO, A. 1958. *Phys. Rev.* **111**, 740.  
MARTIN, D. L. 1960. *Can. J. Phys.* **38**, 17.  
SYDORIAK, S. G. and ROBERTS, T. R. 1957. *Phys. Rev.* **106**, 175.

RECEIVED JUNE 17, 1960.  
DIVISION OF PURE PHYSICS,  
NATIONAL RESEARCH COUNCIL,  
OTTAWA, CANADA.

## OBSERVATIONS ON ESTUARY ICE

P. SCHWERDTFEGER

### INTRODUCTION

During the physical examination of the thermal regime of an ice cover, three types of measurement emerge as being of particular importance. These are the observations of the temperature gradient in the ice, the flux of heat through the ice, and the rate of growth of ice on the lower surface. Standard instruments and techniques allow the continuous recording of the first two, but it is not so

simple to obtain a continuous record of the ice thickness at an unmanned station. It is thus of interest to observe that in river estuaries experiencing a significant tidal movement, newly frozen ice is either fresh or saline, depending on the time or state of the tide, so that under these conditions, a vertical core of ice shows two types of frozen bands, which provide an interesting picture for the study of the rate of growth of ice.

Assuming that freezing has commenced, on the flow of the tide, sea water, under these conditions normally at  $-1.9^{\circ}\text{C}$ , will meet the lower surface of  $0^{\circ}\text{C}$  fresh water ice; before further freezing can take place, the lower surface of the existing ice must be cooled to the freezing point of sea water, so that an abrupt transition between fresh and sea water ice occurs. On the ebb of the tide,  $0^{\circ}\text{C}$  fresh water will move down under freshly frozen sea ice, normally at  $-1.9^{\circ}\text{C}$ ; the warmer fresh water will cause melting of the sea ice, but melting progressively less ice in its path down the estuary as it is cooled and increased in salinity. Thus both a temporary change in temperature gradient, and a permanent salinity gradient are formed in the region of the ice and water boundary, so that on eventual completion of freezing, a continuous transition between sea and fresh water will result. It is of interest to note that this transition layer would only be expected between the lower sea ice and upper fresh water ice boundaries.

#### CHURCHILL RIVER ESTUARY ICE

Several vertical cores were extracted from the Churchill River near Hudson Bay in February 1960. Cores were taken and the thickness of the ice cover was measured at 200- to 400-yard intervals between the river mouth and approximately  $1\frac{1}{2}$  miles upstream, at which point the bands due to the tidal sea water began to disappear. At the mouth, recent fractures had allowed the established ice to drift away, and the new ice formed was only 80 cm thick, far less than the 140 cm found 1 mile upstream. Cores extracted from the thicker ice over a range of more than 200 yards had similarly spaced visible strata; these were thus considered worthy of further study.

#### METHODS AND OBSERVATIONS

A 1 meter long, 3 in. diameter core was prepared for photography by smoothing and polishing with a warm flat iron, and forming two parallel flat surfaces reducing the core thickness to about  $1\frac{1}{2}$  inches. The core was illuminated along its length, so that the only light reaching the camera was scattered at  $90^{\circ}$  from the white bands of sea ice. A photograph taken in this manner is shown in Fig. 1. As Hudson Bay water is of unusually low salinity, near to the estuary, sea water under the ice cover is nearer to  $0^{\circ}\text{C}$  than the usual  $-1.9^{\circ}\text{C}$ , so that the expected transition layer can only be thin. Examination of the photograph does show, however, that the lower white sea ice boundary is more diffuse than the upper.

Because of interest in the crystal structure of the stratified ice, vertical and horizontal sections were planed to 1 mm, mounted on glass slides, and photographed between crossed polaroids. The same samples were then photo-



PLATE I

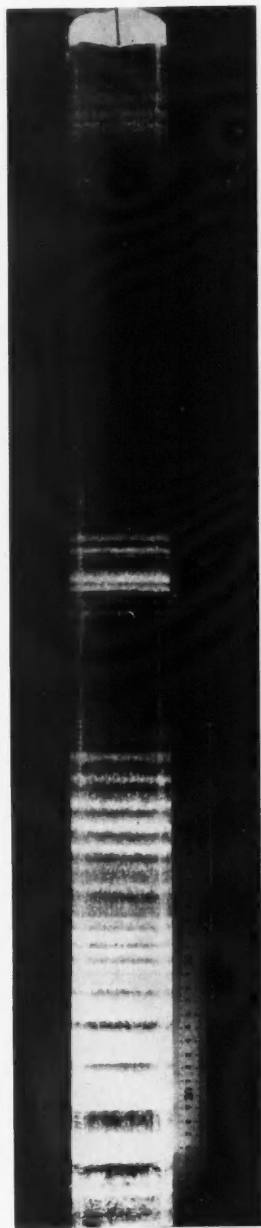
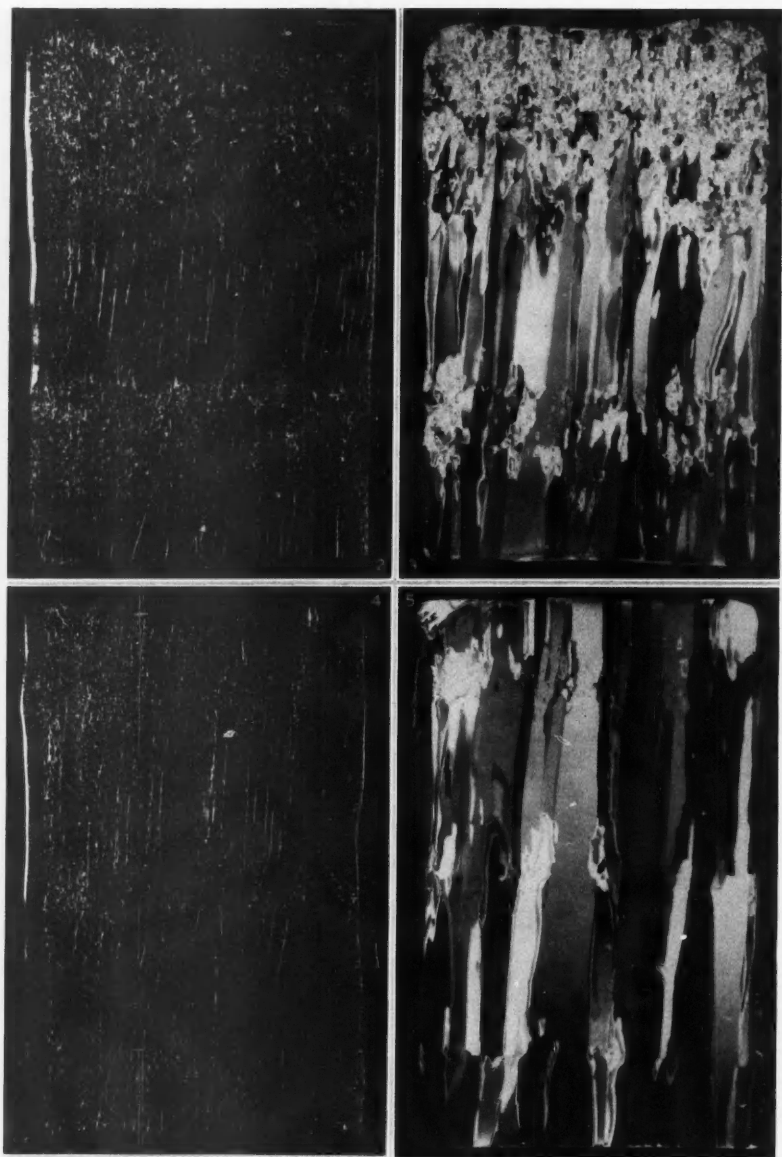


FIG. 1. A vertical core of estuary ice, photographed by scattered light, with cm scale.

PLATE II



FIGS. 2-5. Two thin vertical ice sections photographed by scattered light and through crossed polaroids.

graphed, as was the long core, using side illumination. Figures 2, 3 show a vertical section of the uppermost 4 inches of ice. Figure 2 shows the relatively thick white layers, and the sharp upper and transition type lower boundaries in the case of the lower sea ice stratum. The same sample is shown as viewed through polaroids in Fig. 3; it is interesting to note that the space occupied by sea ice type crystals is considerably less than that taken up by the white sea ice band. Figures 4, 5 show the ice between 8 and 12 inches down from the surface. The sea ice layers, although still  $\frac{1}{2}$  in. thick, are difficult to see on the thin slide of Fig. 4. The interruption in crystal form seen in Fig. 5 is far less than that seen in Fig. 3, the long fresh water crystals tending to continue as if uninterrupted. It thus appears that the technique of establishing the rate of ice growth by means of introducing a thin layer of suitable liquid under the ice cover at regular time intervals would not seriously interrupt the natural structure and growth.

#### RATE OF ICE GROWTH

Under conditions of uniform temperature gradient in the ice, Stefan (1891) has shown that the thickness of ice,  $x$ , after a time  $t$ , may be given by

$$(1) \quad x^2 = \frac{2k}{L\rho} \int_0^t \theta dt'$$

where  $k$  and  $\rho$  are the conductivity and density of the ice, and  $L$  is the latent heat of freezing, all in c.g.s. units. Assuming that the freezing point of the water below the cover is  $0^\circ\text{C}$ ,  $\theta$  is the surface temperature of the ice. Stefan's examination of many arctic records available at the time indicated that equation (1) was of remarkable accuracy even when temperature gradients showed considerable non-uniformity, provided that the surface temperature was not too low.

Differentiation of equation (1) leads to

$$(2) \quad \frac{d(x^2)}{dt} = \frac{2k}{L\rho} \theta.$$

Inserting values for fresh water, since the ratio  $k/L\rho$  will differ very little for low salinity sea water,  $k = 5 \times 10^{-3}$ ,  $L = 80$ ,  $\rho = 0.917$  (c.g.s. units), in (2):

$$(3) \quad d(x^2)/dt = 1.36 \times 10^{-4} \theta.$$

If the square of the ice thickness in centimeters is plotted against the elapsed freezing time in days, the gradient of the resulting curve may be found at any point as  $D$  day  $\text{cm}^{-2}$ , i.e.  $8.64 \times 10^4$  sec  $\text{cm}^{-2}$ . Comparing this statement with equation (3), the following value is obtained for the temperature

$$(4) \quad \theta = 0.085/D.$$

The graph of Fig. 6 was prepared by plotting the squares of the distances of the upper, sharp sea ice boundaries seen in Fig. 1, against time intervals as given by the tides. Since the average time between two high tides is 12

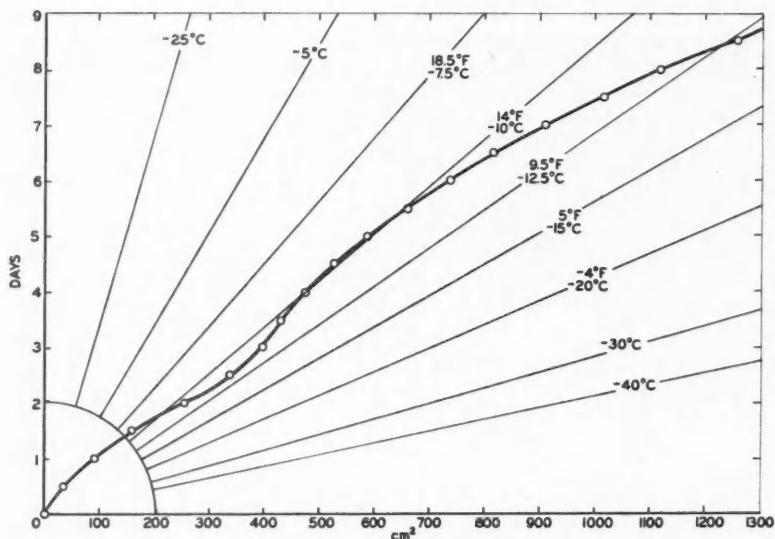


FIG. 6. Cumulative growth of ice.

hours, one fresh water and one neighboring sea ice layer have grown in that time. A number of temperatures in terms of gradients, calculated from (4), are shown indicating that the average surface temperature of the ice during the first 9 days of freezing was about  $-12.5^{\circ}\text{C}$ , which compares well with the average air temperature of  $-14^{\circ}\text{C}$  for the first 10 days of November 1959, recorded in Churchill. The temperature extremes, both high and low, indicated by the graph, are unreliable, since these are usually transient, and do not lead to a uniform temperature gradient in the ice, the requirement for the validity of the above analysis. Despite this restriction, the conductivity / latent heat ratio can be calculated for natural ice, if the rate of ice growth is found on an occasion when the temperature gradient is approximately uniform.

The author wishes to acknowledge his indebtedness to his director of research, Prof. E. R. Pounder, for a critical review of the manuscript. The work was supported by the Defence Research Board under D.D.P. Contract GC-69-900109.

STEFAN, J. 1891. *Ann. Physik Chem. N.F. Bd.* **42**, 269.

RECEIVED JULY 5, 1960.  
ICE RESEARCH PROJECT,  
DEPARTMENT OF PHYSICS,  
MCGILL UNIVERSITY,  
MONTREAL, QUE.

## LETTERS TO THE EDITOR

*Under this heading brief reports of important discoveries in physics may be published. These reports should not exceed 600 words and, for any issue, should be submitted not later than six weeks previous to the first day of the month of issue. No proof will be sent to the authors.*

### Frequency Measurement of Standard Frequency Transmissions<sup>1,2</sup>

Measurements are made at Ottawa, Canada, using N.R.C. caesium-beam frequency resonator as reference standard (with an assumed frequency of 9 192 631 770 c.p.s.). Frequency deviations from nominal are quoted in parts per 10<sup>10</sup>. A negative sign indicates that the frequency is below nominal.

Date, July 1960	MSF, 60 kc/s	GBR, 16 kc/s		WWVB 60 kc/s
		5-hour average*	24-hour average	
1	-154	-152	-154	-69
2	-150	-157	-152	N.M.
3	-149	-154	-152	N.M.
4	N.M.	-155	-154	N.M.
5	-154	-156	-152	-71
6	-148	-148	-152	-70
7	-151	-156	-154	-71
8	-152	-153	-154	-75
9	-159	-168	-164	N.M.
10	-151	-162	-164	N.M.
11	-158	-153	-155	-73
12	-151	-158	-153	-73
13	-151	-156	-158	-71
14	-158	-160	-157	-76
15	-156	-155	-158	-75
16	-157	-157	-159	N.M.
17	-157	-160	-156	N.M.
18	-152	-157	-158	-77
19	-167	-160	-158	-77
20	-159	-154	-157	-77
21	-153	-157	-157	-79
22	-166	-161	-164	-79
23	-159	-160	-158	N.M.
24	-158	-161	-162	N.M.
25	-160	-161	-161	-80
26	N.M.	-162	-162	N.M.
27	N.M.	N.M.	N.M.	-80
28	-162	-160	-159	-154
29	-160	-160	-160	-154
30	-161	-161	-161	N.M.
31	N.M.	N.M.	N.M.	N.M.
Midmonthly mean	-156	-158	-158	
Midmonthly mean of WWV	-150			

NOTE: N.M. no measurement.

\*Time of observations: 01.15 to 03.15 U.T.; 10.00 to 13.00 U.T.

RECEIVED AUGUST 13, 1960.  
DIVISION OF APPLIED PHYSICS,  
NATIONAL RESEARCH COUNCIL,  
OTTAWA, CANADA.

S. N. KALRA

<sup>1</sup>Issued as N.R.C. No. 5955.

<sup>2</sup>Cf. Kalra, S. N. 1959. Can. J. Phys. 37, 1328.



## THE PHYSICAL SOCIETY

MEMBERSHIP of the Society is open to all who are interested in Physics.

FELLOWS pay an Entrance Fee of £1 1s. (\$3.15) and an Annual Subscription of £2 2s. (\$6.00).

STUDENTS: A candidate for Studentship must be between the ages of 18 and 26, and pays an Annual Subscription of 5s. (\$0.75).

MEETINGS: Fellows and Students may attend all Meetings of the Society including the annual Exhibition of Scientific Instruments and Apparatus.

PUBLICATIONS include the *Proceedings of the Physical Society*, published monthly, £12 12s. (\$36.00) per annum, and *Reports on Progress in Physics*, published annually. Volume XXII, 1959, is now available (price £3 3s. (\$9.45)). Members are entitled to receive any of the Publications at a reduced rate.

Further information can be obtained from:

THE PHYSICAL SOCIETY

1, LOWTHER GARDENS, PRINCE CONSORT ROAD  
LONDON, S.W.7, ENGLAND









## NOTES TO CONTRIBUTORS

### *Canadian Journal of Physics*

#### MANUSCRIPTS

**General.**—Manuscripts, in English or French, should be typewritten, double spaced, on paper  $8\frac{1}{2} \times 11$  in. The original and one copy are to be submitted. Tables and captions for the figures should be placed at the end of the manuscript. Every sheet of the manuscript should be numbered. Style, arrangement, spelling, and abbreviations should conform to the usage of recent numbers of this journal. Greek letters or unusual signs should be written plainly or explained by marginal notes. Characters to be set in boldface type should be indicated by a wavy line below each character. Superscripts and subscripts must be legible and carefully placed. Manuscripts and illustrations should be carefully checked before they are submitted. Authors will be charged for unnecessary deviations from the usual format and for changes made in the proof that are considered excessive or unnecessary.

**Abstract.**—An abstract of not more than about 200 words, indicating the scope of the work and the principal findings, is required, except in Notes.

**References.**—References should be listed alphabetically by authors' names, unnumbered, and typed after the text. The form of the citations should be that used in current issues of this journal; in references to papers in periodicals, titles should not be given and only initial page numbers are required. The names of periodicals should be abbreviated in the form given in the most recent *List of Periodicals Abstracted by Chemical Abstracts*. All citations should be checked with the original articles and each one referred to in the text by the authors' names and the year.

**Tables.**—Tables should be numbered in roman numerals and each table referred to in the text. Titles should always be given but should be brief; column headings should be brief and descriptive matter in the tables confined to a minimum. Vertical rules should not be used. Numerous small tables should be avoided.

#### ILLUSTRATIONS

**General.**—All figures (including each figure of the plates) should be numbered consecutively from 1 up, in arabic numerals, and each figure referred to in the text. The author's name, title of the paper, and figure number should be written in the lower left corner of the sheets on which the illustrations appear. Captions should not be written on the illustrations.

**Line drawings.**—Drawings should be carefully made with India ink on white drawing paper, blue tracing linen, or co-ordinate paper ruled in blue only; any co-ordinate lines that are to appear in the reproduction should be ruled in black ink. Paper ruled in green, yellow, or red should not be used. All lines must be of sufficient thickness to reproduce well. Decimal points, periods, and stippled dots must be solid black circles large enough to be reduced if necessary. Letters and numerals should be neatly made, preferably with a stencil (do NOT use typewriting) and be of such size that the smallest lettering will be not less than 1 mm high when the figure is reduced to a suitable size. Many drawings are made too large; originals should not be more than 2 or 3 times the size of the desired reproduction. Whenever possible two or more drawings should be grouped to reduce the number of cuts required. In such groups of drawings, or in large drawings, full use of the space available should be made; the ratio of height to width should conform to that of a journal page ( $4\frac{1}{2} \times 7\frac{1}{2}$  in.), but allowance must be made for the captions. The original drawings and one set of clear copies (e.g. small photographs) are to be submitted.

**Photographs.**—Prints should be made on glossy paper, with strong contrasts. They should be trimmed so that essential features only are shown and mounted carefully, with rubber cement, on white cardboard, with no space between those arranged in groups. In mounting, full use of the space available should be made. Photographs are to be submitted in duplicate; if they are to be reproduced in groups one set should be mounted, the duplicate set unmounted.

#### REPRINTS

A total of 100 reprints of each paper, without covers, are supplied free. Additional reprints, with or without covers, may be purchased at the time of publication.

Charges for reprints are based on the number of printed pages, which may be calculated approximately by multiplying by 0.6 the number of manuscript pages (double-spaced typewritten sheets,  $8\frac{1}{2} \times 11$  in.) and including the space occupied by illustrations. Prices and instructions for ordering reprints are sent out with the galley proof.

## Contents

<i>R. A. Hurd</i> —An electromagnetic diffraction problem involving unidirectionally conducting surfaces - - - - -	1229
<i>L. E. H. Trainor</i> —Considerations on a meson-atomic model of the nucleon	1245
<i>R. E. Bedford</i> —A low temperature standard of total radiation - - -	1256
<i>B. K. Bhattacharyya</i> —A study of auroral motions from all-sky camera records - - - - -	1279
<i>M. A. Thomas and H. L. Welsh</i> —The Raman spectrum of methane - - -	1291
<i>R. A. Armstrong and A. Szabo</i> —Relaxation in ruby - - - - -	1304
<i>B. Muller and M. Bloom</i> —On the study of diffusion using spin echoes - -	1318
<i>C. D. Cox</i> —Bulk photoeffects in inhomogeneous semiconductors - - -	1328
<i>J. S. Kirkaldy</i> —The thermodynamic description of heterogeneous dissipative systems by variational methods. I. A formulation of the principle of minimum rate of entropy production with application to certain stationary heterogeneous convective systems - - - - -	1343
<i>J. S. Kirkaldy</i> —The thermodynamic description of heterogeneous dissipative systems by variational methods. II. A variational principle applicable to non-stationary (unconstrained) heterogeneous dissipative systems -	1356
<i>J. A. Kim and B. W. Currie</i> —Further observations of the horizontal movements of aurora - - - - -	1366
<i>C. B. Benson and A. C. Hollis Hallett</i> —Viscosity measurements in liquid helium II - - - - -	1376
Notes:	
<i>Douglas L. Martin</i> —The specific heat of annealed and cold-worked copper from 0.4° to 1.5° K - - - - -	1390
<i>P. Schwerdtfeger</i> —Observations on estuary ice - - - - -	1391
Letters to the Editor:	
<i>S. N. Kalra</i> —Frequency measurement of standard frequency transmissions	1395

